ROBUST EMPIRICAL BAYES ESTIMATION IN FINITE POPULATION SAMPLING

BY

PARTHASARATHI LAHIRI

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN
PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

1986

to my parents, with regards

ACKNOWLEDGMENTS

I would like to thank Dr. Malay Ghosh for being my thesis advisor and for originally proposing the problem. Without his enormous patience, encouragement and guidance, this work would not have been possible. I would like to thank Dr. Michael A. DeLorenzo, Dr. Kim Sawyer, and Dr. Mark C. K. Yang for serving in my committee. Also, I am grateful to Dr. Ramon C. Littell and Dr. Kenneth M. Portier for serving on my Part C and oral defense committees. Lastly, I would like to thank my parents for just about everything.

TABLE OF CONTENTS

		PAGE
ACKNOWLEDGMENTS		iii
ABSTRACT		vi
CHAPTERS		
ONE	INTRODUCTION	1
	1.1 Literature Review	1
	1.2 The Subject of This Dissertation	4
TWO	ESTIMATION OF MEANS FROM STRATIFIED SAMPLES	7
	2.1 Introduction	7
	2.2 Bayes Estimation	11
	2.3 Empirical Bayes Estimation	12
	2.4 RSL Behavior	19
	2.5 A Numerical Study	34
THREE	ESTIMATION OF MEANS FROM STRATIFIED SAMPLES IN THE PRESENCE OF AUXILIARY INFORMATION	48
	3.1 Introduction	48
	3.2 Bayes Estimation	50
	3.3 Empirical Bayes Estimation	50
	3.4 RSL Behavior	52
FOUR	ESTIMATION OF VARIANCES FROM STRATIFIED SAMPLES	56
	4.1 Introduction	56
	4.2 Bayes Estimation	58
	4.3 Empirical Bayes Estimation	61
	4.4 RSL Behavior	62
FIVE	ESTIMATION OF MEANS FROM TWO-STAGE STRATIFIED SAMPLES	73
	5.1 Introduction	73
	5.2 Bayes Estimation	76
	5.3 Empirical Bayes Estimation	87
	5.4 RSL Behavior	96

SIX SUMMARY AND FUTURE RESEARCH ... 112

6.1 Summary ... 112
6.2 Future Research ... 113

BIBLIOGRAPHY ... 114

BIOGRAPHICAL SKETCH ... 118

Abstract of Dissertation Presented to the Graduate School
of the University of Florida in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy

ROBUST EMPIRICAL BAYES ESTIMATION
IN FINITE POPULATION SAMPLING

By

Parthasarathi Lahiri

August, 1986

Chairman: Dr. Malay Ghosh
Major Department: Statistics

This dissertation concerns simultaneous estimation of several strata means and variances, where each stratum contains a finite number of elements. The results of this dissertation will find application in small area estimation where very few observations are available for one individual area, and estimates of certain parameters (such as the mean or the variance etc.) for an individual area or simultaneous estimates of these parameters for several areas could be improved by utilizing information from similar neighboring sources. For instance, agencies of the federal government have been recently involved in obtaining estimates of items such as population counts, unemployment rates, per-capita income etc. for states or for local government regions.

The observations in each stratum are considered to be realizations from a hypothetical superpopulation depending on some

or more unknown parameters. Priors are imposed on the super-population parameters which make the posterior expectation of the superpopulation mean a linear function of sample observations. This property, referred to as "posterior linearity," holds for many well known parametric families of priors, including but not restricted to, the normal prior. Bayes and different empirical Bayes estimators are found without assuming any specific distributional form of the prior.

Simultaneous estimation of several strata means are considered with and without the presence of auxiliary information. Simultaneous estimation of several strata means is also considered when samples are drawn from each stratum in two stages. Also, we consider simultaneous estimation of several strata variances.

The proposed empirical Bayes estimators are compared with suitable rival estimators on the basis of relative savings loss (RSL). Asymptotic behavior of the RSL is studied under the model and the assumption of posterior linearity without making any further distributional assumptions regarding the superpopulation. This indicates a clear robustness property of the estimation procedure. The proposed empirical Bayes estimators perform very satisfactorily on the basis of this criterion as compared to some of the classical estimators. On the basis of a numerical study, it turns out that the empirical Bayes estimators serve as a compromise between the Bayes and the classical estimators.

CHAPTER ONE
INTRODUCTION

1.1 Literature Review

Early work on model-based inference in survey sampling theory can be found in Cochran's papers (1939, 1946), where the finite population is viewed as a realization from a hypothetical super-population. From a Bayesian perspective, this amounts to putting a prior distribution on the parameters of the finite population. An elegant formulation of the Bayes approach in finite population sampling is due to Ericson (1969a). He put prior on the finite population in two stages. At the first stage, a prior is imposed on the finite population which is known except for some or some unknown parameters. This is equivalent to the classical superpopulation approach of sample survey theory. At the second stage, priors are assumed on the parameter or the vector of parameters of the superpopulation. Ericson found the Bayes estimator of the finite population mean and variance under a normal theory set up. He also considered the estimation problem in the situation of extreme prior vagueness. In such a situation, each element of the finite population is considered to be an outcome from one of several categories. A multinomial super-population with Dirichlet prior imposed on the success

probabilities is assumed. We also considered the problem in the presence of auxiliary information. In certain special cases of assumed diffused prior, we obtained the classical ratio and regression estimators.

In a later paper (1969b), Ericson found an identical estimator for the finite population mean, starting only with the assumption that the posterior expectation of the superpopulation mean is a linear function of the sample observations. This is the so called "posterior linearity" property which is the key assumption in Linear Bayes theory. It was pointed out by Ericson (1969b) that this condition holds for many parametric families used as priors, including but not limited to, the normal prior. In fact, it was shown by Diaconis and Ylvisaker (1979) that "posterior linearity" is preserved for the general exponential family with conjugate prior imposed on the parameter. It was shown in Goldstein (1975b) that the condition holds for Dirichlet process priors (see Ferguson (1973) for the definition of Dirichlet process prior). Yet another example of a parametric family of priors where "posterior linearity" holds is the situation when the prior is taken to be multivariate t. Hartigan (1969) used the linear Bayes idea in the context of prediction theory. Similar ideas have also been used by the actuarists under the heading of "credibility theory" (see for example, Jewell (1974a, b)).

Robbins (1955) introduced the empirical Bayes method in the context of nonparametric estimation of a completely unspecified prior distribution. Subsequently, the method has been explored very successfully in a series of articles by Efron and Morris (1973, 1975, 1977) in a parametric framework. In the Efron-Morris set up, a family of parametric distributions is used as possible priors where one or more parameters of the prior distribution is unknown and needs to be estimated from the data. Morris (1983b) lists a number of areas where empirical Bayes method is useful.

One of the main features of empirical Bayes analysis is to "borrow strength from the ensemble," i.e. use information from similar sources in constructing estimators or predictors in addition to the most directly available source of information. It was pointed out by Ghosh and Meeden (1986) that empirical Bayes analysis could be very suitable in certain estimation problems in finite population sampling. They have applied the method to estimate a single finite population mean. One situation where their method is useful is in repetitive surveys, e.g. in household surveys, crop cutting experiments etc. In each a situation, one has at one's disposal, not only the current data, but also data from similar past experiments. In each cases, past information can be used effectively in conjunction with current information to arrive at suitable estimators.

Fay and Herriot (1979) and Fuller and Harter (1985) have also considered other methods of empirical Bayes estimation different from the ones used by Ghosh and Meeden (1986). However, the methods employed by these authors rest heavily on the normality assumption of the superpopulation.

1.2 The Subject of This Dissertation

In this dissertation, we consider simultaneous estimation of several strata means or variances, where each stratum contains a finite number of elements. The normality assumption is relaxed and different empirical Bayes estimators have been constructed, starting from the assumption of posterior linearity. This is the approach of "Linear Empirical Bayes" introduced by Efron and Morris (1973) in the context of infinite population.

Chapter Two devotes to the simultaneous estimation of several strata means without the presence of auxiliary information. In Chapter Three, we have addressed the same problem when auxiliary information is available. Chapter Four is devoted to the simultaneous estimation of strata variances. Chapter Five considers the estimation of several strata means when samples are drawn from each stratum in two stages. A linear Bayes estimator was found by Ericson (1976) in two stage sampling. But his model and the resulting estimator are different from what is prescribed in this dissertation. Our Bayes estimator is a generalization of

the estimators found by Scott and Smith (1969), and is different from that prescribed in Maier and Sedransk (1985). However, none of these authors addressed the problem of empirical Bayes estimation in simultaneous estimation.

The proposed empirical Bayes estimators are compared with a rival estimator as the basis of relative savings loss (RSL) introduced by Efron and Morris (1973). This is really the proportion of Bayes risk (integrated over the sample and the prior) improvement over the rival estimator that is sacrificed by the use of empirical Bayes estimator instead of using the ideal Bayes estimator.

Ghosh and Meeden (1986) studied the asymptotic behavior of the RSL while estimating a single finite population mean under a normal theory set up. We have been able to show essentially the same asymptotic behavior of the RSL when the normality assumption is replaced by some suitable moment conditions. The exact conditions will be discussed in details in Chapter Two through Chapter Five. We shall show in these chapters a robustness property of our estimation procedures. We have established asymptotic negligibility of the RSL as the number of strata increases to infinity. In certain special cases, we have been able to find the exact rate at which the RSL converges to zero. In the process, we have also proved asymptotic optimality in the sense of Robbins (1955). The conclusion of this dissertation is

that in all cases explored, the proposed empirical Bayes estimators perform very well, at least asymptotically. A numerical study is undertaken in Chapter Two which shows that the proposed empirical Bayes estimator performs well even when only a small number of strata are available.

CHAPTER TWO
ESTIMATION OF MEANS FROM STRATIFIED SAMPLES

2.1 Introduction

Consider a finite population having m strata, the jth stratum containing a finite number of units. Let the jth stratum be denoted by U_j with units labelled $U_{j1},\ldots,U_{jN_j}$. Let X_{ji} denote the characteristic of interest associated with the ith unit in the jth stratum ($i=1,\ldots,N_j; j=1,\ldots,m$). Let s_j denote a typical sample of fixed size n_j taken from the jth stratum. Without loss of generality, let $s_j = (U_{j1},\ldots,U_{jn_j})$ for $j=1,\ldots,m$. A sampling design is the probability of selecting the sample s_j from all possible samples S_j that can be drawn from the jth stratum. Thus, $p(s_j) \geq 0$ for all $s_j \in S_j$ with $\sum_{s_j \in S_j} p(s_j) = 1$. Suppose, without loss of generality, the sample values are denoted by $X_{j1},\ldots,X_{jn_j}$ $(j=1,\ldots,m)$. Let $\underline{X}_j = (X_{j1},\ldots,X_{jn_j})$ and $\bar{X}_j = n_j^{-1}\sum_{i=1}^{n_j} X_{ji}$ $(j=1,\ldots,m)$. We assume no nonsampling error, bias, or nonresponse so that once a sample is drawn, the value of the characteristic is known for sure. In this chapter, we consider the problem of estimation of $\underline{\bar{X}} = (\bar{X}_1,\ldots,\bar{X}_m)$, where

$\bar{Y}_j = N_j^{-1}\Sigma_{i=1}^{N_j} Y_{ji}$ under the loss

$$L(\underset{\sim}{a},\underset{\sim}{\gamma}) = m^{-1}\Sigma_{j=1}^{m}(a_j-\gamma_j)^2. \qquad (\#.1)$$

As pointed out by Ghosh and Meeden (1986), this problem can be identified as a classical prediction problem.

The model we are going to assume is as follows

(i) Conditional on $\underset{\sim}{\theta} = (\theta_1,\ldots,\theta_m)$, let X_{ji}'s ($i=1,\ldots,N_j$, $j=1,\ldots,m$) be mutually independent, and further conditional on θ_j, let X_{ji}'s be iid with distribution depending only on θ_j, and $E(X_{ji}|\theta_j) = \theta_j$, and $V(X_{ji}|\theta_j) = \mu_2(\theta_j)$ ($j=1,\ldots,m$);

(ii) θ_j's are iid with mean μ and variance σ^2. We write $\tau^2 = E\mu_2(\theta_1)$ and assume that $0<\tau^2<\infty$.

Thus, priors are put on the finite population in two stages. The first stage prior stated by (i) is really the so called superpopulation, the finite population being a realization from this superpopulation only. The second stage prior stated by (ii) is the one that a Bayesian would put on the superpopulation parameters. Here, σ^2 is the prior variability while τ^2 can be interpreted as the sampling variability in the context of infinite normal theory set up. Thus, according to our assumptions, observations in any individual stratum are exchangeable while observations from different strata are independent. We could have a slightly more general model with $E(X_{ji}|\theta_j) = g(\theta_j)$, where $g(\theta_j)$ is not necessarily a one-one function of θ_j, but we sacrifice this generality for simplicity. Also, our assumptions are met in most

of the important applications, e.g. when X_{j1}'s have distributions belonging to the one parameter general exponential family with conjugate priors imposed on the parameters.

The basic assumption of this chapter is that

$$E(\theta_j|\underline{X}_j) = \sum_{i=1}^{\ell_j} a_{ji} X_{ji} + b_j \quad (j=1,\ldots,m), \tag{2.1.2}$$

where the a_{ji}'s and b_j's are constants not depending on $\underline{X}_j$. In view of the fact that X_{ji}'s are conditionally iid given θ_j, it follows from the work of Goldstein (1975b) that (2.1.2) leads to

$$E(\theta_j|\underline{X}_j) = a_j \bar{X}_j + b_j \quad (j=1,\ldots,m), \tag{2.1.3}$$

where a_j's are some constants. Equation (2.1.2) or (2.1.3) is the so called posterior linearity property mentioned in the Introduction of this dissertation. Henceforth, we shall assume (2.1.3) rather than (2.1.2). Apart from the priors noted in Ericson (1969b), Goldstein (1975b), and Diaconis and Ylvisaker (1979), the "posterior linearity" property holds in the following situation.

Conditional on $\underline{\theta}$ and h, X_{ji}'s are mutually independent with X_{ji}'s iid $N(\theta_j, h^{-1})$. Also, conditional on h, θ_j's are iid $N(\mu, (\lambda h)^{-1})$, marginally h is a gamma random variable with scale parameter a and shape parameter β, i.e., h has pdf

$$g_{a,\beta}(h) = \exp(-ah)h^{\beta-1}a^{\beta}/\Gamma(\beta), \; h > 0, \; a > 0 \text{ and } \beta > 0, \tag{2.1.4}$$

Writing $\underline{X}_j = (X_{j1},\ldots,X_{jn_j})$, under the model (2.1.4), it can be shown that

$$E(\theta_j|\underline{X}_1,\ldots,\underline{X}_m) = E\{E(\theta_j|\underline{X}_1,\ldots,\underline{X}_m,h)|\underline{X}_1,\ldots,\underline{X}_m\}$$

$$= \mu + \frac{n_j}{\lambda+n_j}(\bar{X}_j - \mu). \qquad (2.1.5)$$

This is clearly of the form (2.1.3) with $a_j = a_j/(\lambda + a_j)$ and $b_j = \mu\lambda(\lambda + n_j)^{-1}$. In this situation, the joint marginal distribution of the $\underline{X}_j$'s is multivariate t. Unlike our earlier assumption, in this case the θ_j's are only iid (normal) conditional on h. But, the results of this section based on the iid assumptions of the θ_j's also hold under (2.1.4) by using the simple device of first conditioning on h.

From Ericson (1969b) or Hartigan (1969), it follows that $a_j = \sigma^2/(\sigma^2+\tau^2 n_j^{-1}) = a_j/(N+n_j) = 1-B_j$, where $N = \tau^2/\sigma^2$ and $B_j = N(N+n_j)^{-1}$, while $b_j = B_j\mu$ $(j=1,\ldots,m)$. Thus, (2.1.3) leads to

$$E(\theta_j|\underline{X}_j) = (1-B_j)\bar{X}_j + B_j\mu. \qquad (2.1.6)$$

Thus, the posterior mean of the superpopulation mean can be interpreted as an weighted average of the sample mean and the prior mean.

Ghosh and Meeden (1986) considered empirical Bayes estimation of a single finite population mean assuming a normal superpopulation model. Their estimators generalize automatically when

one estimates a vector of stratum means. However, such estimators enjoy a certain robustness property in the sense that they can be motivated even without the normality assumptions. In Section 2.2, we achieve this by deriving the Bayes estimators of several strata means under our model and the posterior linearity assumption (2.1.3). In Section 2.3, different empirical Bayes estimators are obtained by replacing prior parameters in the Bayes estimator obtained in Section 2.2 by their respective estimators. When samples of equal sizes are drawn from different strata, as special cases of the proposed empirical Bayes estimators, one gets finite population analogue of the James-Stein estimators, plus rule versions of James-Stein estimators, and the estimators of Lindley and Smith (1972). In Section 2.4, we show the asymptotic optimality under the existence of certain superpopulation moments.

2.2 Bayes Estimation

Under the loss $L(a,\chi)$, eg. (2.1.1), the Bayes estimator of χ is $a_B = (a_B^{(1)} \ldots a_B^{(m)})$, where

$$
\begin{aligned}
& a_B^{(j)} \\
&= E(\bar{Y}_j \,|\, \underline{X}_j) \\
&= N_j^{-1}\left[\sum_{i=1}^{n_j} X_{ji} + \sum_{i=n_j+1}^{N_j} E(X_{ji} \,|\, \underline{X}_j)\right] \\
&= N_j^{-1}\left[\sum_{i=1}^{n_j} X_{ji} + \sum_{i=n_j+1}^{N_j} E\{E(X_{ji} \,|\, \underline{X}_j, \theta_j) \,|\, \underline{X}_j\}\right]
\end{aligned}
$$

$$= N_j^{-1}[n_j\bar{X}_j + (N_j - n_j)\, E(\theta_j \mid \underline{X}_j)]$$

$$= N_j^{-1}[n_j\bar{X}_j + (N_j - n_j)(B_j\mu + (1-B_j)\bar{X}_j)]$$

$$= (1-f_jB_j)\bar{X}_j + f_jB_j\mu$$

$$= \bar{X}_j - f_jB_j(\bar{X}_j - \mu), \qquad (2.2.1)$$

$f_j = N_j^{-1}(N_j - n_j)$ denoting the finite population correction factor for the j^{th} stratum.

Thus, the Bayes estimator of the jth stratum mean has two nice interpretations. It can be interpreted as a weighted average of the sample mean and the prior mean. Also, it is a shrinkage estimator, shrinking the sample mean towards the prior mean. The factor f_j is due to the fact that samples are drawn from a finite population.

2.3 Empirical Bayes Estimation

In an empirical Bayes framework, at least one of the parameters μ and M in (2.2.1) is unknown and needs to be estimated from the data. First we estimate M.

Let $n_T = \sum_{j=1}^{m} n_j$, $\bar{\bar{X}} = \sum_{j=1}^{m} n_j\bar{X}_j/n_T$, $SSA = \sum_{j=1}^{m} n_j(\bar{X}_j - \bar{\bar{X}})^2$, $MSB = SSB/(m-1)$, $SSW = \sum_{j=1}^{m}\sum_{i=1}^{n_j}(X_{ji} - \bar{X}_j)^2$, and $MSW = (n_T - m)^{-1}SSW$. Here, we tacitly assume that $n_T > m$ which holds if, for example, $n_j \geq 2$ for all j. We now prove a result which motivates

estimation of M. The result was proved by Ghosh and Meeden (1986) under the normality assumption.

Theorem 2.3.1. $E(MSW) = \tau^2$ and $E(MSB) = \tau^2 + g\sigma^2(m-1)^{-1}$, where $g = g(n_1,\ldots,n_m) = n_T - (\sum_{j=1}^{m} n_j^2)/n_T$.

Let us first state and prove the following Lemma giving the means and variances of quadratic forms of independent random variables under the assumption of finite fourth moments. This Lemma will be used repeatedly in proving the above theorem and many other theorems and results in this chapter and subsequent chapters.

Lemma 2.3.1. Let $W_1,\ldots,W_k$ be independent random variables with $E(W_r) = 0$, $E(W_r^2) = \mu_{2(r)}$ and $E(W_r^4) = \mu_{4(r)}$ where $\mu_{4(r)} < \infty$ for all $r = 1,\ldots,k$. Then, for any symmetric $k\times k$ matrix $\underset{\sim}{A} = ((a_{rr'}))$,

$$E(\underset{\sim}{W}\underset{\sim}{A}\underset{\sim}{W}^T) = \sum_{r=1}^{k} a_{rr}\mu_{2(r)}, \tag{2.3.1}$$

$$V(\underset{\sim}{W}\underset{\sim}{A}\underset{\sim}{W}^T) = \sum_{r=1}^{k} a_{rr}^2(\mu_{4(r)} - \mu_{2(r)}^2) + 2\sum_{1\le r<r'\le k} a_{rr'}^2\mu_{2(r)}\mu_{2(r')}, \tag{2.3.2}$$

where $\underset{\sim}{W} = (W_1,\ldots,W_k)$.

Proof of Lemma 2.3.1 Using the independence of W_r's and the fact that $E(W_r) = 0$ for all r, it follows that

$$E(\underset{\sim}{W}\underset{\sim}{A}\underset{\sim}{W}^T) = \sum_{r=1}^{k} a_{rr}E(W_r^2) + \sum_{r\neq r'} a_{rr'}E(W_r)E(W_{r'}) = \sum_{r=1}^{k} a_{rr}\mu_{2(r)}.$$

This proves (2.3.1). Next we compute,

$$E(\underline{W}\underline{A}\underline{W}^T)^2$$

$$=E\Big[\sum_{r=1}^{k} a_{rr} W_r^2 + \sum_{r\neq r'} a_{rr'} W_r W_{r'}\Big]^2$$

$$=E\Big[\sum_{r=1}^{k} a_{rr}^2 W_r^4 + \sum_{r\neq r'} a_{rr'} a_{r'r'} W_r^2 W_{r'}^2 + 2\sum_{r\neq r'} a_{rr'}^2 W_r^2 W_{r'}^2\Big]$$

$$=\sum_{r=1}^{k} a_{rr}^2 \mu_4(r) + \sum_{r\neq r'} a_{rr} a_{r'r'} \mu_2(r)\mu_2(r') + 2\sum_{r\neq r'} a_{rr'}^2 \mu_2(r)\mu_2(r')$$

$$=\sum_{r=1}^{k} a_{rr}^2 (\mu_4(r) - \mu_2^2(r)) + \Big\{\sum_{r=1}^{k} a_{rr} \mu_2(r)\Big\}^2 + 2\sum_{r\neq r'} a_{rr'}^2 \mu_2(r)\mu_2(r'). \tag{2.3.3}$$

Combining (2.3.1) and (2.3.3), one gets (2.3.2).

Proof of Theorem 2.3.1. Write $Z_{ji} = X_{ji} - \theta_j$ so that $E(Z_{ji}|\theta_j) = 0$ and $V(Z_{ji}|\theta_j) = \mu_2(\theta_j)$. Recall that $E\mu_2(\theta_j) = \tau^2$ for all j. Define $\bar{Z}_j = n_j^{-1} \sum_{i=1}^{n_j} Z_{ji}$ and $\bar{\bar{Z}} = \sum_{j=1}^{m} n_j \bar{Z}_j / n_T$.

Note that conditional on θ_j, Z_{ji}'s are independently distributed. Now, writing $\underline{Z}_j = (Z_{j1}, \ldots, Z_{jn_j})$, I_{n_j} as the identity matrix of order n_j, and $\underline{1}_{n_j}$ as a n_j-component row vector with all elements equal to 1, one gets

$$\sum_{i=1}^{n_j} (Z_{ji} - \bar{Z}_j)^2 = \underline{Z}_j (I_{n_j} - n_j^{-1} \underline{1}_{n_j}^T \underline{1}_{n_j}) \underline{Z}_j^T.$$

Next, applying (2.3.1),

$$E(SSW) = \sum_{j=1}^{m} E\{E[\sum_{i=1}^{n_j} (Z_{ji} - \bar{Z}_j)^2 | \theta_j]\}$$

$$= \sum_{j=1}^{m} E\Big[\sum_{i=1}^{n_j} (1 - n_j^{-1}) \mu_2(\theta_j)\Big] = (n_T - m)\tau^2. \tag{2.3.4}$$

This proves E(MSW) = τ^2. Next we compute

$$E(SSA) = \Sigma_{j=1}^{m} n_j E[(\bar{z}_j - \bar{\bar{z}})^2 + (\theta_j - \bar{\theta})^2 + 2(\bar{z}_j - \bar{\bar{z}})(\theta_j - \bar{\theta})], \quad (2.3.5)$$

where $\bar{\theta} = \Sigma_{j=1}^{m} n_j \theta_j / n_T$. Write $\underset{\sim}{\theta} = (\theta_1, \ldots, \theta_m)$, $\underset{\sim}{D} = \mathrm{diag}(n_1, \ldots, n_m)$ and $\underset{\sim}{n} = (n_1, \ldots, n_m)$. Then, $E(\bar{z}_j - \bar{\bar{z}} \mid \underset{\sim}{\theta}) = 0$. Applying (2.3.1) once again,

$$E[\Sigma_{j=1}^{m} n_j (\theta_j - \bar{\theta})^2] = E[\Sigma_{j=1}^{m} n_j [(\theta_j - u) - (\bar{\theta} - u)]^2]$$

$$= E[(\underset{\sim}{\theta} - u\underset{\sim}{1}_m)^T (\underset{\sim}{D} - \frac{1}{n_T} \underset{\sim}{n}\underset{\sim}{n}^T)(\underset{\sim}{\theta} - u\underset{\sim}{1}_m)]$$

$$= \Sigma_{j=1}^{m} (n_j - \frac{n_j^2}{n_T}) a^2 = g\sigma^2, \quad (2.3.6)$$

$$E[\Sigma_{j=1}^{m} n_j (\bar{z}_j - \bar{\bar{z}})^2 \mid \underset{\sim}{\theta}] = E[(\bar{z}_1 \ldots \bar{z}_m)(\underset{\sim}{D} - \frac{1}{n_T} \underset{\sim}{n}\underset{\sim}{n}^T)(\bar{z}_1 \ldots \bar{z}_m)^T \mid \underset{\sim}{\theta}]$$

$$= \Sigma_{j=1}^{m} (n_j - \frac{n_j^2}{n_T}) \tau^2 n_j^{-1} = \tau^2 (m-1). \quad (2.3.7)$$

Thus, from (2.3.5) - (2.3.7),

$$E(SSA) = g\sigma^2 + (m-1)\tau^2. \quad (2.3.8)$$

The proof of the theorem is complete.

From Theorem 2.3.1, a sensible estimator of $M^{-1} = \sigma^2/\tau^2$ is given by

$$\hat{M}^{-1} = \max\{0, (\frac{(m-1)MSB}{(n-g)MSW} - 1)(n-1)g^{-1}\}. \quad (2.3.9)$$

The corresponding estimator of B_j is $\hat{B}_j = (1+n_j\hat{M}^{-1})^{-1}$. These estimators are the same as the ones given in Ghosh and Maunden

(1986). It is pointed out in the latter that the multiplier $(m-1)/(m-3)$ of MSE/MSW makes $\hat{M}^{-1}$ identical with finite population analogue of the pice rule version of Jmeam-Stein estimator in the balanced case, i.e. when $n_1 = \ldots = n_m$. In case one has knowledge about the distributional form of the superpopulation, one can possibly go for Bayes empirical Bayes (Hierarchial bayes) approach where priors are assumed on the hyperparameters (in our case μ, τ^2, σ^2). But, even in the case of normality of the superpopulation, the resulting estimator turns out to be in the form of a complicated integral whose value can only be found by numerical integration.

It remains to estimate μ. Write $\underset{\sim}{X} = (\underset{\sim}{X}_1, \ldots, \underset{\sim}{X}_m)$. Then $E(\underset{\sim}{X}) = \mu \underset{\sim}{1}_{n_T}$, and $V(\underset{\sim}{X})$ is a block diagonal matrix with its jth diagonal block equal to $\tau^2 \underset{\sim}{1}_{n_j} \underset{\sim}{1}_{n_j}^T + \sigma^2 \underset{\sim}{1}_{n_j}^T \underset{\sim}{1}_{n_j}$ $(j=1,\ldots,m)$. Then, if τ^2 and σ^2 are known, the BLUE of μ is obtained by minimizing

$$(\underset{\sim}{X} - \mu \underset{\sim}{1}_{n_T})[V(\underset{\sim}{X})]^{-1}(\underset{\sim}{X} - \mu \underset{\sim}{1}_{n_T})^T$$

$$= \Sigma_{j=1}^m (\underset{\sim}{X}_j - \mu \underset{\sim}{1}_{n_j})(\tau^2 \underset{\sim}{1}_{n_j} \underset{\sim}{1}_{n_j}^T + \sigma^2 \underset{\sim}{1}_{n_j}^T \underset{\sim}{1}_{n_j})^{-1}(\underset{\sim}{X}_j - \mu \underset{\sim}{1}_{n_j})^T \qquad (2.3.10)$$

with respect to μ. If the underlying distributions are assumed to be normal, the BLUE is identical with the MLE of μ, and then it follows from Ghosh and Meeden (1986) that the resulting estimator of μ is

$$\mu_* = \Sigma_{j=1}^m (1-B_j)\bar{X}_j / \Sigma_{j=1}^m (1-B_j). \qquad (2.3.11)$$

When the B_j's are unknown, μ is estimated by

$$\hat{\mu} = \Sigma_{j=1}^{m}(1-\hat{B}_j)\bar{X}_j / \Sigma_{j=1}^{m}(1-\hat{B}_j) \quad \text{if } \hat{M}^{-1} \neq 0$$

$$= m^{-1}\Sigma_{j=1}^{m}\bar{X}_j \quad \text{if } \hat{M}^{-1} = 0. \qquad (2.3.12)$$

Now, from (2.2.1), an empirical Bayes estimator δ_{EB} of χ is given by $\delta_{EB} = (e_{EB}^{(1)}, \ldots, e_{EB}^{(m)})$, where

$$e_{EB}^{(j)} = \bar{X}_j - f_j\hat{B}_j(\bar{X}_j - \hat{\mu}), \qquad j=1,\ldots,m. \qquad (2.3.13)$$

Note that the estimator δ_{EB} is a vector valued extension of the estimator a_{EB} of Ghosh and Meeden (1986). In case when μ is unknown, but M and accordingly the B_j's are known, one estimates μ by μ_w instead of $\hat{\mu}$, and the resulting empirical Bayes estimator of χ is $\delta_{EB}^{0} = (e_{EB}^{0(1)} \ldots e_{EB}^{0(m)})$, where

$$e_{EB}^{0(j)} = \bar{X}_j - f_j B_j(\bar{X}_j - \mu_w). \qquad j=1,\ldots,m. \qquad (2.3.14)$$

In case when $n_1 = \ldots = n_m = n$, $g = n(m-1)$ so that the $\hat{M}^{-1}$ given in (2.3.9) reduces to

$$\hat{M}^{-1} = \max\{0, \left(\frac{(m-1)MSB}{(m-3)MSW} - 1\right) n^{-1}\}. \qquad (2.3.15)$$

The corresponding $B = (1 + nM^{-1})^{-1}$ is then estimated by $\hat{B} = (1+n\hat{M}^{-1})^{-1}$ which is the plus rule version of the James-Stein shrinker. In this case, μ is estimated by $\bar{\bar{X}} = m^{-1}\Sigma_{j=1}^{m}\bar{X}_j$, and one gets the corresponding empirical Bayes estimator

$\underset{\sim}{e}^{*}_{EB} = (e^{*(1)}_{EB},\ldots,e^{*(m)}_{EB})$, where

$$e^{*(j)}_{EB} = \bar{X}_j - f_j \hat{B}(\bar{X}_j - \bar{\bar{X}}), \quad j=1,\ldots,m. \tag{2.3.16}$$

In this case, if instead one estimates M^{-1} by $\hat{M}^{-1} = \left(\frac{(m-1)MSB}{(m-3)MSW} - 1\right)n^{-1}$ so that the resulting estimator of B is $\hat{B} = \frac{(m-3)SSW}{m(n-1)SSB}$ and the corresponding empirical Bayes estimator is $\underset{\sim}{e}_{JS} = (e^{(1)}_{JS},\ldots,e^{(m)}_{JS})$, where

$$e^{(j)}_{JS} = \bar{X}_j - f_j \hat{B}(\bar{X}_j - \bar{\bar{X}}), \tag{2.3.17}$$

which equals Lindley's modification of the Jeams-Stein estimator in the infinite population case, i.e., when $f_j = 1$ for all $j=1,\ldots,m$. We should also note that if $n_1=\ldots=n_m$ and M and accordingly B is known, $\underset{\sim}{\chi}$ is estimated by $\underset{\sim}{e}^{**}_{EB} = (e^{**(1)}_{EB},\ldots,e^{**(m)}_{EB})$, where

$$e^{**(j)}_{EB} = \bar{X}_j - f_j B(\bar{X}_j - \bar{\bar{X}}), \quad j=1,\ldots,m \tag{2.3.18}$$

which is the finite population analogue of the estimator of Lindley and Smith (1972). This estimator was generated by Lindley and Smith in a normal theory set up.

Efron and Morris (1973) have advocated the use of empirical Bayes estimators in an infinite population context starting with linear Bayes estimators. They have named the resulting empirical Bayes estimators "linear empirical Bayes estimators," but have not

put forward any concrete proposal of how to construct such estimators. Robbins (1983) has also considered linear empirical Bayes estimators again in an infinite population framework. But his basic model as well as the methods of estimation of prior parameters are different from ours, and accordingly the resulting estimators are different.

2.4 RSL Behavior

Next, we compare the Bayes risk (integrated over both the parameter and the sample space) performance of $\underset{\sim}{\delta}_{EB}$ with the classical estimator $\underset{\sim}{\delta}_0 = (\bar{X}_1,\ldots,\bar{X}_m)$ of $\underset{\sim}{\mu}$. Let $r(\xi,\delta)$ denote the Bayes risk of an estimator δ under the prior ξ mentioned in Section 2.1. Rather than comparing $r(\xi,\underset{\sim}{\delta}_{EB})$ directly with $r(\xi,\underset{\sim}{\delta}_0)$, we compare a normalized version of the Bayes risk improvement of $\underset{\sim}{\delta}_{EB}$ over $\underset{\sim}{\delta}_0$. With this end, we introduce the notion of relative savings loss (RSL) as given in Efron and Morris (1973) or in Ghosh and Meeden (1986). The RSL of $\underset{\sim}{\delta}_{EB}$ with respect to an arbitrary estimator $\underset{\sim}{\delta}_0$ under a prior ξ is given by

$$RSL(\xi,\underset{\sim}{\delta}_{EB},\underset{\sim}{\delta}_0) = \frac{r(\xi,\underset{\sim}{\delta}_{EB})-r(\xi,\underset{\sim}{\delta}_B)}{r(\xi,\underset{\sim}{\delta}_0)-r(\xi,\underset{\sim}{\delta}_B)} = 1 - \frac{r(\xi,\underset{\sim}{\delta}_0)-r(\xi,\underset{\sim}{\delta}_{EB})}{r(\xi,\underset{\sim}{\delta}_0)-r(\xi,\underset{\sim}{\delta}_B)}. \qquad (2.4.1)$$

Thus, $RSL(\xi;\underset{\sim}{\delta}_{EB},\underset{\sim}{\delta}_0)$ denotes the possible Bayes risk improvement over $\underset{\sim}{\delta}_0$ that is sacrificed by the use of empirical Bayes estimator $\underset{\sim}{\delta}_{EB}$ instead of using the ideal Bayes estimator $\underset{\sim}{\delta}_B$ when

the prior is ξ. The RSL expression given in (2.4.1) is useful when $r(\xi,\underline{e}_0) > r(\xi,\underline{e}_{\xi\underline{e}})$.

The following lemma is a simple extension of Lemma 3 of Ghosh and Meeden (1986). We shall use this Lemma repeatedly in deriving different RSL results.

Lemma 2.4.1 For any estimator $\underline{e}$,

$$r(\xi,\underline{e})-r(\xi,\underline{e}_{\xi}) = m^{-1}\sum_{j=1}^{m} E\{e_{\underline{e}}^{(j)}-e_{\xi}^{(j)}\}^2, \tag{2.4.2}$$

where E denotes expectation over the prior distribution ξ given in Section 2.1 as well as the sample.

Proof of Lemma 2.4.1 Using the fact that

$$E[(Y_j-e_{\underline{e}}^{(j)})|\underline{X}_j] = E(Y_j|\underline{X}_j) - e_{\xi}^{(j)} = 0$$

and $e_{\underline{e}}^{(j)}-e^{(j)}$ is a function of $\underline{X}$, one gets

$$\begin{aligned} r(\xi,\underline{e}) &= m^{-1}\sum_{j=1}^{m} E\{e^{(j)}-Y_j\}^2 \\ &= m^{-1}\sum_{j=1}^{m} E\{e^{(j)}-e_{\xi}^{(j)}+e_{\xi}^{(j)}-Y_j\}^2 \\ &= r(\xi,\underline{e}_{\xi})+m^{-1}\sum_{j=1}^{m} E\{e^{(j)}-e_{\xi}^{(j)}\}^2. \end{aligned} \tag{2.4.3}$$

This completes the proof of the Lemma.

Using the above Lemma, one gets the following theorem which is a simple extension of Theorem 1 of Ghosh and Meeden (1986).

Theorem 2.4.1

$$\mathrm{RSL}(\hat{\theta}; \underset{\sim}{\theta}, \underset{\sim}{\theta}_0)$$

$$= \Big[\sum_{j=1}^{m} f_j^2 \mathrm{E}[(\hat{\theta}_j - \theta_j)(\bar{X}_j - \mu) - \hat{\theta}_j(\hat{\mu} - \mu)]\Big]^2 / \sum_{j=1}^{m} f_j^2 \theta_j^2 \mathrm{E}(\bar{X}_j - \mu)^2. \tag{2.4.4}$$

To study the asymptotic behavior of the RSL expression given in (2.4.4), we first prove that $\hat{M}$ is a consistent estimator of M. The result was proved by Ghosh and Meeden (1986) under the normality assumption. Theorem 2.4.2 of the present paper dispenses with the normality assumption, but invokes instead certain moment assumptions on the θ_j's. These assumptions, however, are fairly mild and are easily verifiable.

Theorem 2.4.2. Assume that (i) $E\mu_4(\theta_j) < \infty$, where $\mu_4(\theta_j) = E(X_{j1}^4 | \theta_j)$, (ii) $E(\theta_1^4) < \infty$, (iii) $\inf_{j \geq 1} n_j \geq 2$ and (iv) $\sup_{j \geq 1} n_j = K < \infty$. Then, $\hat{M} \xrightarrow{P} M$ as $m \to \infty$.

Remark 1. Assumptions (iii) and (iv) were also needed in Ghosh and Meeden (1986). Note that assumptions (i) and (ii) are usually compatible, and in many instances, one is implied by the other. For example, if conditional on θ_j, the X_{ji}'s are iid Poisson (θ_j), $\mu_4(\theta_j)$'s are quadratic functions of θ_j's so that (ii) implies (i). The same phenomenon holds in the binomial case. In the normal case, typically one assumes that $\mu_2(\theta_j) = \tau^2$ so that $\mu_4(\theta_j) = 3\tau^4$, which is a constant.

Proof of Theorem 2.4.2 Throughout this section, we shall use the notation C for a generic constant, not depending on m.

$$V(SSW) = \sum_{j=1}^{m} V[\Sigma_{i=1}^{n_j}(Z_{ji} - \bar{Z}_j)^2]. \tag{2.4.5}$$

Using (2.3.1),

$$E\{\Sigma_{i=1}^{n_j}(Z_{ji} - \bar{Z}_j)^2 | \theta_j\} = (n_j - 1)\mu_2(\theta_j)$$

so that

$$V[E\{\Sigma_{i=1}^{n_j}(Z_{ji} - \bar{Z}_j)^2 | \theta_j\}] = (n_j - 1)^2 V(\mu_2(\theta_1)) = (n_j - 1)^2[E\mu_2^2(\theta_1) - \nu^4]. \tag{2.4.6}$$

Again, applying (2.3.2),

$$\begin{aligned} &V(\Sigma_{i=1}^{n_j}(Z_{ji} - \bar{Z}_j)^2 | \theta_j) \\ &= \Sigma_{i=1}^{n_j}(1 - n_j^{-1})^2(\mu_4(\theta_j) - \mu_2^2(\theta_j)) + \Sigma_{i \neq j} n_j^{-2}\mu_2^2(\theta_j) \\ &= n_j(1 - n_j^{-1})^2(\mu_4(\theta_j) - \mu_2^2(\theta_j)) + 2(n_j - 1)n_j^{-1}\mu_2^2(\theta_j), \\ &= n_j^{-1}(n_j - 1)[\mu_4(\theta_j)(n_j - 1) - (n_j - 3)\mu_2^2(\theta_j)] \end{aligned}$$

so that

$$\begin{aligned} &E[V(\Sigma_{i=1}^{n_j}(Z_{ji} - \bar{Z}_j)^2 | \theta_j] \\ &= n_j^{-1}(n_j - 1)E[(n_j - 1)\mu_4(\theta_1) - (n_j - 3)\mu_2^2(\theta_1)]. \end{aligned} \tag{2.4.7}$$

From (2.4.5) - (2.4.7), it follows that

$$V(SSW) = \{\Sigma_{j=1}^{m}(n_j-1)^2 n_j^{-1}\}E\mu_4(\theta_1) + \Sigma_{j=1}^{m}\{(n_j-1)^2-(e_j-1)(n_j-3)e_j^{-1}\}E\mu_2^2(\theta_j) - \Sigma_{j=1}^{m}(e_j-1)^2\tau^4. \tag{2.4.8}$$

Since $E\mu_2^2(\theta_1) \leq E\mu_4(\theta_1) < \infty$, it follows from (2.4.8) that

$$V(SSW) < Cm. \tag{2.4.9}$$

Hence, since $n_T \geq 2m$,

$$V(MSW) < Cm\,(n_T-m)^{-2} < Cm^{-1}. \tag{2.4.10}$$

Also, from Theorem 2.3.1, $E(MSW) = \tau^2$. Thus, using (2.4.10)

$$MSW \xrightarrow{P} \tau^2. \tag{2.4.11}$$

If we can show now that $V(MSB) \to 0$ as $m \to \infty$, then in view of (2.4.11), $(MSB-E(MSB))/MSW \xrightarrow{P} 0$. Use (2.4.11) and the bound $g(m-1)^{-1} \leq n_T(m-1)^{-1} \leq 2K$ to get $E(MSB) = O(1)$ so that $MSB = O_p(1)$. Hence, we have

$$(m-1)(m-3)^{-1}MSB/MSW - E(MSB)/\tau^2 \xrightarrow{P} 0. \tag{2.4.12}$$

Again $E(MSB)/\tau^2 = 1 + gM^{-1}(m-1)^{-1}$, and we have shown already that $g(m-1)^{-1} = O(1)$. Recalling the definition of M^{-1} in (2.3.9), one gets the result.

It remains only to show that V(MSB) → 0 as n → ∞. Using the inequality $V(U_1+U_2+U_3) \le 3[V(U_1)+V(U_2)+V(U_3)]$, and using the representation of SSB as in (2.3.5), one gets

$$V(SSB)$$

$$\le 3[V\{\Sigma_{j=1}^{m} n_j(\bar{Z}_j - \bar{\bar{Z}})^2\} + V\{\Sigma_{j=1}^{m} n_j(\theta_j - \bar{\theta})^2\} + V\{\Sigma_{j=1}^{m} n_j(Z_j - \bar{\bar{Z}})(\theta_j - \bar{\theta})\}]. \qquad (2.4.13)$$

Now,

$$V[\Sigma_{j=1}^{m} n_j(\bar{Z}_j - \bar{\bar{Z}})(\theta_j - \bar{\theta})]$$

$$= V[\Sigma_{j=1}^{m} n_j \bar{Z}_j(\theta_j - \bar{\theta})]$$

$$= E[V[\Sigma_{j=1}^{m} n_j \bar{Z}_j(\theta_j - \bar{\theta}) \mid \underset{\sim}{\theta}]]$$

$$= E[\Sigma_{j=1}^{m} n_j \mu_2(\theta_j)(\theta_j - \bar{\theta})^2]$$

$$= n_1 E[\mu_2(\theta_1)(\theta_1 - \bar{\theta})^2]$$

$$\le n_1 E^{1/2}[\mu_2^2(\theta_1)] E^{1/2}(\theta_1 - \bar{\theta})^4 \le mK E^{1/2}[\mu_4(\theta_1)] E^{1/2}(\theta_1 - \bar{\theta})^4. \qquad (2.4.14)$$

Next, applying (2.3.2), one gets

$$V[\Sigma_{j=1}^{m} n_j(\theta_j - \bar{\theta})^2]$$

$$= V[\Sigma_{j=1}^{m} n_j\{(\theta_j - \mu) - (\bar{\theta} - \mu)\}^2]$$

$$= \Sigma_{j=1}^{m} n_j^2\left(1-\frac{n_j}{n_T}\right)^2\left\{E[(\theta_j-\mu)^4]-[E(\theta_j-\mu)^2]^2\right\}$$

$$+ 2\,\Sigma_{1\le j\ne j'\le m}\left(n_j^2 n_{j'}^2/n_T^2\right) E(\theta_j-\mu)^2\, E(\theta_{j'}-\mu)^2$$

$$\le C[mE(\theta_1-\mu)^4 + m^4]. \tag{2.4.15}$$

Finally, applying (2.3.2) and (2.3.7), one gets

$$V[\Sigma_{j=1}^{m} n_j(\bar{Z}_j-\bar{\bar{Z}})^2]$$

$$= E[\Sigma_{j=1}^{m} n_j^2\left(1-\frac{n_j}{n_T}\right)^2\{E(\bar{Z}_j^4|\theta_j) - [E(\bar{Z}_j^2|\theta_j)]^2\}]$$

$$+ 2\,\Sigma_{j\ne j'} n_j^2 n_{j'}^2 n_T^{-2}\, E(\bar{Z}_j^2|\theta_j)\, E(\bar{Z}_{j'}^2|\theta_{j'}). \tag{2.4.16}$$

But,

$$E(\bar{Z}_j^2|\theta_j) = V(\bar{Z}_j|\theta_j) = n_j^{-1}\mu_2(\theta_j)$$

and

$$E(\bar{Z}_j^4|\theta_j) = n_j^{-3}\mu_4(\theta_j) + 6n_j^{-3}(n_j-1)\mu_2^2(\theta_j).$$

Hence, from (2.4.18) using $E\mu_2^2(\theta_1) \le E\mu_4(\theta_1)$, one gets

$$V[\Sigma_{j=1}^{m} n_j(\bar{Z}_j-\bar{\bar{Z}})^2] \le CmE\mu_4(\theta_1). \tag{2.4.17}$$

From (2.4.13) - (2.4.15) and (2.4.17), it follows that

$V(SSB) = O(m)$, so that $V(MSB) = O(m^{-1})$. This completes the proof of Theorem 2.4.2.

Since $\hat{M} \xrightarrow{P} M$ as $m \to \infty$, from Lemma 2 of Ghosh and Meeden (1986), one gets

$$\max_{1\le j\le m} \left|\hat{\theta}_j - \theta_j\right| \xrightarrow{P} 0 \text{ as } m \to \infty. \tag{2.4.18}$$

We next state the main result of this section which establishes the asymptotic convergence of the RSL expression given in (2.4.4) to zero. In the course of proving this result, we prove also the "asymptotic optimality" (cf Robbins (1955)) of the empirical Bayes estimator δ_{EB} defined in (2.3.13).

Theorem 2.4.3. Under the assumptions (i) - (iv) of Theorem 2.4.2,

$$RSL(\tilde{\pi}, \delta_{EB}, \delta_\pi) \to 0 \text{ as } m \to \infty. \tag{2.4.19}$$

Proof: First, we find that using (iii) and (iv) of Theorem 2.4.2

$$\begin{aligned} & m^{-1}\textstyle\sum_{j=1}^m f_j^2 n_j^2 E(\bar{X}_j - \mu)^2 \\ &= m^{-1}\textstyle\sum_{j=1}^m f_j^2 n_j^2 (\sigma^2 + \tau^2 n_j^{-1}) \\ &> (\sigma^2 + \frac{1}{K}\tau^2) M^2 (M+K)^{-2} \inf_{1\le j\le m} f_j^2. \end{aligned} \tag{2.4.20}$$

However,

$$f_j = 1 - n_j N_j^{-1} > 1 - n_j (n_j+1)^{-1} = (n_j+1)^{-1} > (K+1)^{-1}. \tag{2.4.21}$$

Combining (2.4.20) and (2.4.21), one gets

$$m^{-1}\textstyle\sum_{j=1}^m f_j^2 n_j^2 E(\bar{X}_j - \mu)^2 > (\sigma^2 + \frac{1}{K}\tau^2) M^2 (M+K)^{-2} (1+K)^{-2}. \tag{2.4.22}$$

In view of (2.4.22), it remains to show that

$$m^{-1}\sum_{j=1}^{m} E[(\hat{B}_j - B_j)(\bar{X}_j - \mu) - \hat{B}_j(\hat{\mu} - \mu)]^2 \to 0 \text{ as } m \to \infty. \tag{2.4.23}$$

Also, note that (2.4.23) proves the "asymptotic optimality" of $\hat{\xi}_{EB}$ in the sense of Robbins (1955).

To prove (2.4.23), first observe that from (2.4.18), $1_{[\hat{B}_j \neq 1]}|\hat{B}_j - B_j| \xrightarrow{P} 0$. Again, since

$$E[m^{-1}\sum_{j=1}^{m}(\bar{X}_j - \mu)^2]$$

$$= m^{-1}\sum_{j=1}^{m}\{\sigma^2 + \tau^2\sigma_j^{-1}\} \leq \sigma^2 + \frac{1}{2}\tau^2, \tag{2.4.24}$$

it follows that $m^{-1}\sum_{j=1}^{m}(\bar{X}_j - \mu)^2 = O_p(1)$. Also, using arguments similar to (A.5) - (A.9) of Ghosh and Meeden (1986), it follows that $\hat{\mu} - \mu \xrightarrow{P} 0$. Thus,

$$V_m \overset{\text{def}}{=} m^{-1}\sum_{j=1}^{m}[(\hat{B}_j - B_j)(\bar{X}_j - \mu) - \hat{B}_j(\hat{\mu}-\mu)]^2$$

$$\leq 2[1_{[\hat{B}_j \neq 1]}|\hat{B}_j - B_j|^2 m^{-1}\sum_{j=1}^{m}(\bar{X}_j - \mu)^2] + 2(\hat{\mu} - \mu)^2$$

$$\xrightarrow{P} 0 \text{ as } m \to \infty. \tag{2.4.25}$$

To prove the uniform integrability of V_m, first use the inequality

$$V_m \leq 2m^{-1}\sum_{j=1}^{m}[(\hat{B}_j - B_j)^2(\bar{X}_j - \mu)^2 + \hat{B}_j^2(\hat{\mu}-\mu)^2]$$

$$\leq 2[m^{-1}\sum_{j=1}^{m}(\bar{X}_j - \mu)^2 + (\hat{\mu}-\mu)^2]. \tag{2.4.26}$$

But,

$$E[m^{-1}\Sigma_{j=1}^{m}(\bar{X}_j - \mu)^2]^2 \leq m^{-1}\Sigma_{j=1}^{m}E(\bar{X}_j - \mu)^4$$

$$\leq m^{-1}8\ \Sigma_{j=1}^{m}\{E(\bar{X}_j - \theta_j)^4 + E(\theta_j - \mu)^4\}$$

$$= 8[m^{-1}\Sigma_{j=1}^{m}\ E(\bar{Z}_j^4) + E(\theta_1 - \mu)^4].$$

Using the inequality

$$E(\bar{Z}_j^4) \leq (\sigma_j^{-3} + 6n_j^{-2})\ E\ \mu_4(\theta_2), \tag{2.4.28}$$

it follows from (2.4.28) that

$$\sigma^{-4}\ \Sigma_{j=1}^{m}E\ (\bar{Z}_j^4) \leq \frac{13}{8}\ E\ \mu_4(\theta_1). \tag{2.4.29}$$

Also, since $E(\theta_1^4) < \infty$,

$$E(\theta_1 - \mu)^4 \leq 8[E(\theta_1^4) + \mu^4] \leq 16\ E\theta_1^4 < \infty.$$

This proves the uniform integrability of the first term in the rhs of (2.4.27). Thus, in order to complete the proof of the theorem, it remains only to show that $E(\hat{\mu} - \mu)^4$ remains uniformly bounded by a constant not depending on m for all $m > 1$. But, as in Ghosh and Meeden (1986)

$$(\hat{\mu} - \mu)^2 \leq (\tfrac{1}{2}K + 1)\ m^{-1}\Sigma_{j=1}^{m}(\bar{X}_j - \mu)^2. \tag{2.4.30}$$

The proof of the theorem is now complete from (2.4.27) - (2.4.29). In some special cases, more accurate asymptotic behavior of

RSL $(\xi; \overset{0}{\xi}_{EB}, \xi_{\infty})$ can be specified. First suppose M is known, but μ is unknown. In this case, the empirical Bayes estimator is given in (2.3.14). Thus, one gets

$$\begin{aligned} \text{RSL}\,(\xi; \overset{0}{\xi}_{EB}, \xi_{\infty}) &= \Sigma_{j=1}^{m} f_j^2 B_j^2 E(\mu_* - \mu)^2 / \Sigma_{j=1}^{m} f_j^2 B_j^2 E(\bar{X}_j - \mu)^2 \\ &= m^{-1} \Sigma_{j=1}^{m} f_j^2 B_j^2 E(\mu_* - \mu)^2 / m^{-1} \Sigma_{j=1}^{m} f_j^2 B_j^2 E(\bar{X}_j - \mu)^2. \end{aligned} \tag{2.4.31}$$

In view of (2.4.22), we need only show that numerator of (2.4.31) is of order m^{-1}. Now, $f_j^2 B_j^2 \leq 1$ and

$$E(\mu_* - \mu)^2$$

$$= \Sigma_{j=1}^{m} (1-B_j)^2 (\sigma^2 + \tau^2 v_j^{-1}) / \{\Sigma_{j=1}^{m} (1-B_j)\}^2$$

$$\leq (\sigma^2 + \tfrac{1}{2}\tau^2) / \Sigma_{j=1}^{m} (1-B_j)$$

$$\leq (2m)^{-1} (M+2)(\sigma^2 + \tfrac{1}{2}\tau^2). \tag{2.4.32}$$

From (2.4.31), (2.4.22) and (2.4.32), it follows that

$$\text{RSL}(\xi; \overset{0}{\xi}_{EB}, \xi_{\infty}) = O(m^{-1}). \tag{2.4.33}$$

Next, we consider the case $n_1 = \ldots = n_m = n$. First consider the case when M is known, but μ is unknown, so that ξ is estimated by $\overset{**}{\xi}_{EB}$. It is easy to check in this case that

$$\text{RSL}(\xi; \overset{**}{\xi}_{EB}, \xi_{\infty})$$

$$= \{B^2 \Sigma_{j=1}^{m} f_j^2 E(\bar{X} - \mu)^2\} / \{B^2 \Sigma_{j=1}^{m} f_j^2 E(\bar{X}_j - \mu)^2\} = m^{-1}. \tag{2.4.34}$$

Finally, in this section, we consider the case when $n_1=\ldots=n_m=n$, but M and μ are both unknown. In this case χ is estimated by either $\underset{\sim}{\hat{\xi}}_{EB}^{*}$ or $\underset{\sim}{\hat{\xi}}_{JS}$. The RSL expressions are given respectively by

$$RSL(\underset{\sim}{\hat{\xi}}_{EB}^{*},\underset{\sim}{\xi}_{0})$$

$$= \Sigma_{j=1}^{m} f_j^2 E[\hat{B}(\bar{X}_j - \bar{\bar{X}}) - B(\bar{X}_j - \mu)]^2 / [\Sigma_{j=1}^{m} f_j^2 B^2 E(\bar{X}_j - \mu)^2], \quad (2.4.35)$$

$$RSL(\underset{\sim}{\hat{\xi}}_{JS},\underset{\sim}{\xi}_{0})$$

$$= \Sigma_{j=1}^{m} f_j^2 E[\hat{\hat{B}}(\bar{X}_j - \bar{\bar{X}}) - B(\bar{X}_j - \mu)]^2 / [\Sigma_{j=1}^{m} f_j^2 B^2 E(\bar{X}_j - \mu)^2]. \quad (2.4.36)$$

In the remainder of this Section, we assume that $X_1,\ldots,X_m$ are iid. First we compute $RSL(\underset{\sim}{\xi},\underset{\sim}{\hat{\xi}}_{JS},\underset{\sim}{\xi}_{0})$. It is immediately evident that

$$\Sigma_{j=1}^{m} f_j^2 B^2 E(\bar{X}_j - \mu)^2 = Br^2 \Sigma_{j=1}^{m} f_j^2. \quad (2.4.37)$$

Also, using the fact that $(\hat{\hat{B}}-B)(\bar{X}_j - \bar{\bar{X}})$ has the same distribution for every $j = 1,\ldots,m$ and $(\hat{\hat{B}}-B)(\bar{X}_j - \bar{\bar{X}})(\bar{\bar{X}} - \mu)$ has the same distribution for every $j = 1,\ldots,m$, it follows that

$$E[\hat{\hat{B}}(\bar{X}_j - \bar{\bar{X}}) - B(\bar{X}_j - \mu)]^2$$

$$= E[(\hat{\hat{B}}-B)(\bar{X}_j - \bar{\bar{X}}) - B(\bar{\bar{X}} - \mu)]^2$$

$$= E[(\hat{\hat{B}}-B)^2(\bar{X}_j - \bar{\bar{X}})^2 + B^2(\bar{\bar{X}}-\mu)^2 - 2(\hat{\hat{B}}-B)B(\bar{\bar{X}}-\mu)(\bar{X}_j - \bar{\bar{X}})]$$

$$= E[(\hat{\hat{B}}-B)^2 m^{-1} \Sigma_{j=1}^{m} (\bar{X}_j - \bar{\bar{X}})^2 + (mn)^{-1} Br^2 - 2(\hat{\hat{B}}-B)B(\bar{\bar{X}}-\mu) m^{-1} \Sigma_{j=1}^{m} (\bar{X}_j - \bar{\bar{X}})]$$

$$= (mn)^{-1}\ \{E[(\hat{\bar{\xi}}-\bar{\xi})^2 SSE] + B\tau^2\}. \qquad (2.4.38)$$

Now, writing $c = (m-3)/(m(n-1))$ and using Theorem 2.3.1, it follows that

$$E[(\hat{\bar{\xi}}-\bar{\xi})^2 SSE]$$

$$= c^2 E[(SSW)^2/SSE] - 2cBE(SSW) + B^2 E(SSE)$$

$$= c^2 E[(SSW)^2/SSE] - 2(m-3)B\tau^2 + (m-1)B\tau^2$$

$$= c^2 E[(SSW)^2/SSE] - (m-5)B\tau^2. \qquad (2.4.39)$$

This leads to the following theorem.

Theorem 2.4.3. Under the assumed model with the extra condition that $X_1, \ldots, X_m$ are iid,

$$RSL(\xi, \xi_{JS}, \xi_0)$$

$$= (mn)^{-1}\{[(m-3)/(m(n-1))]^2 E[(SSW)^2/SSE](B\tau^2)^{-1} - (m-6)\}. \qquad (2.4.40)$$

Suppose, now

$$E[(SSW)^2/SSE] = E(SSW)^2/E(SSE) + O(1). \qquad (2.4.41)$$

We shall show that under the assumption (2.4.41),

$$RSL(\xi, \xi_{JS}, \xi_0) = O(m^{-1}). \qquad (2.4.42)$$

When $n_1=\ldots=n_m=n$, one gets from (2.4.8),

$$V(SSW) = m\left[(n-1)^2n^{-1}E\mu_4(\theta_1)+(n-1)(n^2-2n+3)n^{-1}E\mu_2^2(\theta_1)-(n-1)^2\tau^4\right], \tag{2.4.43}$$

so that

$$E(SSW)^2 = V(SSW) + (E(SSW))^2$$

$$= m\left[(n-1)^2n^{-1}E\mu_4(\theta_1)+(n-1)(n^2-2n+3)n^{-1}E\mu_2^2(\theta_1)\right]+m(m-1)(n-1)^2\tau^4. \tag{2.4.44}$$

We have already seen in (2.4.39) that $E(SSB) = (m-1)\tau^2/B$. Hence, from (2.4.44),

$$E(SSW)^2/E(SSB)$$

$$= a(n-1)^2B\tau^2 + O(1). \tag{2.4.45}$$

Hence, from (2.4.39), (2.4.41) and (2.4.45), one gets

$$E\left[(\hat{\bar{\theta}}-\bar{\theta})^2SSB\right]$$

$$= \left(a^{-1}(m-3)^2-(m-5)\right)B\tau^2 + O(1) = \left(-1+9m^{-1}\right)B\tau^2 + O(1). \tag{2.4.46}$$

Equation (2.4.42) follows now from (2.4.36) - (2.4.38) and (2.4.46).

Before giving examples where (2.4.41) holds, we see an important consequence of (2.4.46). Calculations similar to

(2-4.38) lead to

$$E[\hat{a}(\bar{X}_j - \bar{\bar{X}})^2 - a(\bar{X}_j - \mu)]^2$$

$$= (mn)^{-1}[E[(\hat{b}-b)^2 SSB] + 8\tau^2]. \qquad (2.4.47)$$

Since $(\hat{b}-b)^2 \leq (\hat{\hat{b}}-b)^2$, it follows from (2.4.38), (2.4.39), (2.4.46), and (2.4.47) that

$$RSL(\hat{\xi}_{EB}^{*}, \xi_0) \leq RSL(\hat{\xi}_{JS}, \xi_0) = O(m^{-1}). \qquad (2.4.48)$$

Next, to give examples where (2.4.41) holds, first consider the normal set up of Ghosh and Meeden (1986). In this case, SSW and SSB are independently distributed with $SSW \sim \tau^2 \chi^2_{m(n-1)}$ and $SSB \sim \tau^2 b^{-1} \chi^2_{m-1}$. Accordingly, simple calculations yield

$$E[(SSW)^2/(SSB)] = 8\tau^2\, m(n-1)[m(n-1)+2]/(m-3) \qquad (2.4.49)$$

for $m \geq 4$, while for $m \geq 2$,

$$E(SSW)^2/E(SSB) = 8\tau^2\, m(n-1)[m(n-1)+2]/(m-1), \qquad (2.4.50)$$

so that (2.4.41) holds. Now from (2.4.40), and (2.4.49), one gets

$$RSL(\hat{\xi}, \xi_{JS}, \xi_0) = \frac{3n-1-\frac{6}{m}}{mn(n-1)}. \qquad (2.4.51)$$

Similar calculations are given in Rao (1976).

As a second example where (2.4.41) holds, consider the model given in (2.1.4). In this case, conditional on b, SSW and SSB are independently distributed with $SSW \sim b^{-1}\chi^2_{m(n-1)}$ and

$SSE \sim (\mathbb{h})^{-1}\chi^2_{m-1}$, where $B = \lambda/(\lambda+n)$. Accordingly,

$$E[(SSW)^2/SSE] = EE[(SSW)^2/SSE|h]$$

$$= E(h^{-1})B\ m(n-1)[m(n-1)+2)/(m-3). \qquad (2.4.52)$$

Similarly,

$$E(SSW)^2/E(SSB) = E(h^{-1})\ m(n-1)[m(n-1)+2)/(m-1). \qquad (2.4.53)$$

Thus, if the shape parameter B of the gamma distribution given in (2.1.4) exceeds 1, so that $E(h^{-1}) < \infty$, it follows from (2.4.52) and (2.4.53) that (2.4.41) holds.

2.5 A Numerical Study

In this section, we shall study the Bayes risk performances of different estimators proposed in Section 2.2 and Section 2.3 under different conditions. Robustness of the empirical Bayes estimator is studied under the violation of the assumptions of Section 2.1.

Let μ_j be the true mean of the jth stratum for $j = 1,\ldots,m$. In Sections 2.1-2.4, we have assumed $\mu_1 = \ldots = \mu_m$. In this case, the Bayes estimator of $\underset{\sim}{\mu}$ is $\underset{\sim}{e}_B$. But, if the means are not all equal to μ, $\underset{\sim}{e}_B$ is no longer the Bayes estimator. In such a situation, the Bayes estimator would be $\underset{\sim}{e}_{BT} = (e_{BT}^{(1)},\ldots,e_{BT}^{(m)})$, where

$$e_{BT}^{(j)} = \bar{X}_j - f_j B_j(\bar{X}_j - \mu_j), \text{ for } j = 1,\ldots,m. \qquad (2.5.1)$$

We first compute Bayes risks of different estimators. Write $\bar{X}_{cj} = \sum_{i=n_j+1}^{N_j} X_{ji}/(N_j-n_j)$. Then,

$$r(\xi,\underline{\underline{\delta}}_0) = m^{-1}\sum_{j=1}^{m} E\{\bar{X}_j - \gamma_j\}^2$$

$$= m^{-1}\sum_{j=1}^{m} E\{\bar{X}_j - \frac{n_j\bar{X}_j+(N_j-n_j)\bar{X}_{cj}}{N_j}\}^2$$

$$= m^{-1}\sum_{j=1}^{m} f_j^2 E\{\bar{X}_j-\bar{X}_{cj}\}^2$$

$$= m^{-1}\sum_{j=1}^{m} f_j^2 E\{(\bar{X}_j-\mu_j) - (\bar{X}_{cj}-\mu_j)\}^2$$

$$= m^{-1}\sum_{j=1}^{m} f_j^2\{E(\bar{X}_j-\mu_j)^2 + E(\bar{X}_{cj}-\mu_j)^2 - 2E(\bar{X}_j-\mu_j)(\bar{X}_{cj}-\mu_j)\}$$

$$= m^{-1}\sum_{j=1}^{m} f_j^2\{(\sigma^2+\frac{\tau^2}{n_j}) + (\sigma^2+\frac{\tau^2}{N_j-n_j}) - 2\sigma^2\}$$

$$= m^{-1}\sum_{j=1}^{m} f_j u_j^{-1}\tau^2. \qquad (2.5.2)$$

$$r(\xi,\underline{\underline{\delta}}_{BT}) = m^{-1}\sum_{j=1}^{m} E[\pi_{BT}^{(j)} - \gamma_j]^2$$

$$= m^{-1}\sum_{j=1}^{m} f_j^2 E\{(1-B_j)(\bar{X}_j-\mu_j) - (\bar{X}_{cj}-\mu_j)\}^2$$

$$= m^{-1}\sum_{j=1}^{m} f_j^2\{(1-B_j)^2(\sigma^2+\frac{\tau^2}{n_j}) + (\sigma^2+\frac{\tau^2}{N_j-n_j}) - 2(1-B_j)\sigma^2\}$$

$$= m^{-1}\sum_{j=1}^{m} f_j^2\{B_j u^2 + \frac{\tau^2}{h_j-n_j}\}, \qquad (2.5.3)$$

Using (2.4.2),

$$r(\xi,\underline{\underline{\delta}}_B) = m^{-1}\sum_{j=1}^{m} E[\pi_B^{(j)} - \pi_{BT}^{(j)}]^2 + r(\xi,\underline{\underline{\delta}}_{BT})$$

$$= m^{-1}\sum_{j=1}^{m} f_j^2 B_j^2(u_j-\mu)^2 + r(\xi,\underline{\underline{\delta}}_{BT}) \qquad (2.5.4)$$

and

$$r(\xi,\delta_{EB})$$

$$= m^{-1}\sum_{j=1}^{m} E\{e_{EB}^{(j)} - e_{BT}^{(j)}\}^2 + r(\xi,\delta_{BT})$$

$$= m^{-1}\sum_{j=1}^{m} f_j^2 E[(\hat{B}_j\hat{\mu}-B_j\mu_j) - (\hat{B}_j-B_j)\bar{X}_j]^2 + r(\xi,\delta_{BT}) \qquad (2.5.4)$$

In our study, we have considered 5 strata, i.e. $m = 5$, and have assumed that a 10% sample has been taken from each of the 5 strata. Also, for our study, we have fixed $\tau^2 = 1$ and $\mu = 0$. We have considered both the balanced and unbalanced situations. Our numerical findings are summarized in Tables 2.1 to 2.3 and Figures 2.1 to 2.5.

First consider an unbalanced case where $n_1 = 1$, $n_2 = 2$, $n_3 = 3$, $n_4 = 5$, $n_5 = 10$. Then, (2.5.2)-(2.5.5) lead to

$$r(\xi,\delta_0) = .384, \qquad (2.5.6)$$

$$r(\xi,\delta_{BT}) = .81[(m^{-1}\Sigma_{j=1}^{m}B_j)^2 + .0474], \qquad (2.5.7)$$

$$r(\xi,\delta_B) = .81m^{-1}\Sigma_{j=1}^{m}B_j^2(\mu_j-\mu)^2 + r(\xi,\delta_{BT}), \qquad (2.5.8)$$

$$r(\xi,\delta_{EB}) = .81m^{-1}\Sigma_{j=1}^{m}E[(\hat{B}_j\hat{\mu}-B_j\mu_j) - (\hat{B}_j-B_j)\bar{X}_j]^2 + r(\xi,\delta_{BT}). \qquad (2.5.9)$$

Note that we have been able to find out the Bayes risk of all the estimators except δ_{ES} explicitly in terms of the parameters of the prior distribution without making any assumption regarding the

distributional form of the ptiot. To evoluatm thm expectation in (2.5.9), we havm adopted Monte Carlo simulation. Fot gmnmtating tandom numbmts ftom normal dmviatm, we have usmd thm RANNOR gmnmtating fonction of SAS libmaty. Wm havm calculatmd thm avmtagm of thm mxpmmsion within thm patmntheses in (2.5.9) on thm completion of 1000 mxpmtimmnts. Wm shall discuss thm following 3 cmsms:

Case 1 $\mu_1 = \ldots = \mu_5 = \mu = 0$,

Case 2 $\mu_1 = \ldots = \mu_5 = 0$, but $\mu = 5$,

Case 3 $\mu_1 = 0$, $\mu_2 = 2$, $\mu_3 = 4$, $\mu_4 = 6$, $\mu_5 = 3$, and $\mu = 0$.

Table 2.1 and Figure 2.1 give thm Bayes risks of thm thmm mstimatmts $\underline{x}_B$, $\underline{x}_0$, and $\underline{x}_{EB}$ in Cmsm 1. Notm that, in this situation, $\underline{x}_B$ is idmntical to $\underline{x}_{BT}$. Thm mmpiticml Bayes mstimatot $\underline{x}_{EB}$ pmtfotms in bmtwmmn thm idmal Bayes mstimatot and thm clmssicml sample mmam mstimatot in tmtms of Bayes tisk. The gmmatmt thm bmlief wm havm on thm ptiot (mffmctmd by smallmt values of σ^2), thm bmttmt is thm pmtfotmantm of $\underline{x}_B$ and $\underline{x}_{EB}$ as compmtmd to $\underline{x}_0$. Also, in this situation, wm havm calculatmd thm RSL givmn in (2.4.4) and PCTIMP dmfinmd by

$$\text{PCTIMP} = \frac{r(\xi,\underline{x}_0) - r(\xi,\underline{x}_{EB})}{r(\xi,\underline{x}_0)} \times 100. \qquad (2.5.10)$$

PCTIMP is tmally thm pmtcmntagm of Bayes tisk improvmmmnt if onm

uses δ_{EB} instead of δ_0. It turns out that for smaller values of σ^2, we are getting impressive PCTIMP and RSL values. Figure 2.4 plots the PCTIMP as a function of σ^{-1}.

Table 2.2 gives the risks in Case 2. The risks for δ_0, δ_{BT}, δ_{EB} in this case are exactly the same as those in Case 1, indicating that a misspecification of the common prior mean has no impact on these estimators. But, the performance of δ_B is much worse than that of δ_0 or δ_{EB}. The graph is shown in Figure 2.2.

In Case 3, the exchangeability assumption is lacking. The risk analysis has been carried out in Table 2.3 and Figure 2.3. The violation of this assumption has great effect on δ_B, but the effect is very mild on δ_{EB} which is still performing better than δ_0.

Overall, the Bayes estimator δ_B is very sensitive to inaccurate specification of the prior mean. The empirical Bayes estimator as well as the classical estimator are much less vulnerable to this. On the other hand, if there is indeed exchangeability, the empirical Bayes estimator outperforms the classical estimator, and does substantially better than the subjective Bayes estimator for misspecified priors. Thus, for exchangeable priors, the empirical Bayes estimator enjoys a great degree of robustness.

Table 2.1 The Risk, the Percent Risk Improvement, and the RSL for the Case 1.

c	$r(\xi, \hat{\xi}_{EB})$	$r(\xi, \hat{\xi}_{B})$	PCTIMP	RSL
0.10	0.1656	0.0462	56.8750	0.3535
0.30	0.1971	0.0933	48.6719	0.3534
0.50	0.2381	0.1505	37.9948	0.3752
0.60	0.2574	0.1767	32.9688	0.3893
0.70	0.2738	0.2004	28.6719	0.4003
0.80	0.2885	0.2213	24.8698	0.4130
0.90	0.3014	0.2396	21.5104	0.4280
0.95	0.3070	0.2479	20.0521	0.4342
1.00	0.3125	0.2556	18.6198	0.4431
1.30	0.3365	0.2920	12.3698	0.4837
1.50	0.3474	0.3091	9.5313	0.5113
2.00	0.3635	0.3365	5.3385	0.5684
2.50	0.3717	0.3516	3.2031	0.6204
3.00	0.3762	0.3607	2.0313	0.6652

Table 2.2 The Risk, the Percent Risk Improvement, and the RSL for Case 2.

σ	$r(\xi,\xi_{ET})$	$r(\xi,\xi_{EB})$	$r(\xi,\xi_{B})$	PCTIMP	RSL
0.10	0.0462	0.1656	18.7472	56.8750	0.3535
0.30	0.0933	0.1971	11.9790	48.6719	0.3571
0.50	0.1505	0.2381	6.9956	37.9948	0.3752
0.60	0.1767	0.2574	5.3795	32.9688	0.3893
0.80	0.2213	0.2885	3.2847	24.8698	0.4130
0.85	0.2308	0.2952	2.9272	23.1250	0.4204
0.90	0.2396	0.3014	2.6178	21.5104	0.4280
0.95	0.2479	0.3070	2.3498	20.0521	0.4342
1.00	0.2556	0.3125	2.1172	18.6198	0.4431
1.30	0.2920	0.3365	1.2308	12.3698	0.4837
1.50	0.3091	0.3474	0.9282	9.5313	0.5113
2.00	0.3365	0.3635	0.5841	5.3385	0.5684
2.50	0.3516	0.3717	0.4662	3.2031	0.6204
3.00	0.3607	0.3762	0.4200	2.0313	0.6652

Table 2.3 The Risk, the Percent Risk Improvement, and the RSL for Case 3.

σ	$r(\xi,\xi_{BT})$	$r(\xi,\xi_{EB})$	$r(\xi,\xi_{\hat{\theta}})$	PCTIMP	RSL
0.3	0.0933	0.3757	5.3434	2.1614	0.9714
0.5	0.1505	0.3760	2.5559	2.0833	0.9657
0.8	0.2213	0.3794	1.0072	1.1979	0.9717
1.0	0.2556	0.3816	0.6637	0.6250	0.9813
1.5	0.3091	0.3837	0.4152	0.0781	0.9960
2.0	0.3356	0.3831	0.3739	0.2344	0.9810
2.5	0.3516	0.3827	0.3678	0.3385	0.9599

Next, we consider a balanced situation when $n_1 = \ldots = n_5 = 4$. We fix $\tau^2 = \sigma^2 = 1$, and $\mu_1 = \ldots = \mu_5 = 0$. Thus, (2.5.2)-(2.5.5) lead to

$$r(\xi,\xi_0) = .225;$$

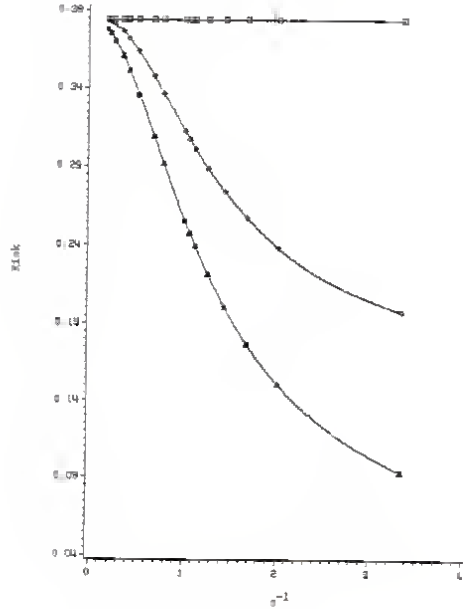
$$r(\xi,\xi_{BT}) = .0278,$$

$$r(\xi,\xi_{\hat{\theta}}) = .0324\mu^2 + .0278.$$

$$r(\xi,\xi_{EB}) = .81a^{-1}\sum_{j=1}^{a} E[\bar{X}_{\cdot\cdot} - (\hat{B}-B)\bar{X}_{j\cdot}]^2 + .0278$$

$$= .81 \times .0302 + .0278 = .0523.$$

Thus, we find that the risks of the estimators ξ_{BT}, ξ_{EB}, and ξ_{0} remain unaffected by a change in the value of μ. The risk of ξ_{B} is affected very much by such a change. In fact, it is a quadratic function of μ. Figure 2.5 demonstrates the behavior of the Bayes risk as a function of μ.



□ denotes risk for $\underline{s}_0$

Δ or * denotes risk for $\underline{s}_\beta$

◇ denotes risk for $\underline{s}_{\underline{g}\beta}$

Figure 2.1 Risk Analysis for Case 1.

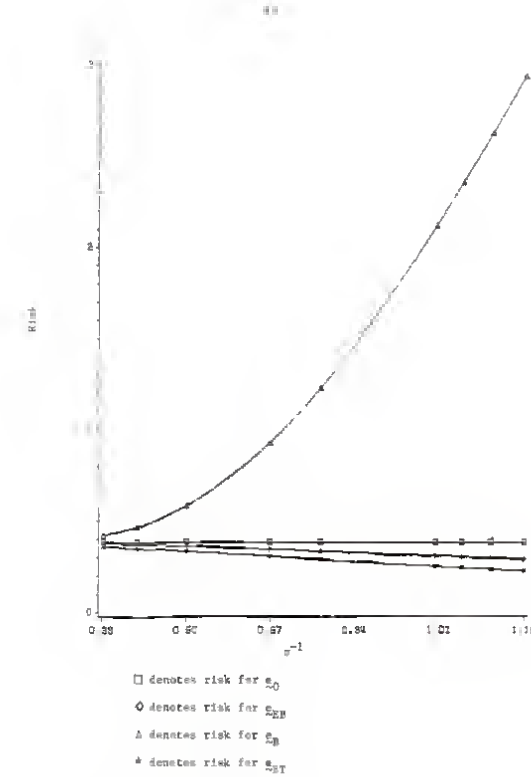


Figure 2.2 Risk Analysis for Case 2.

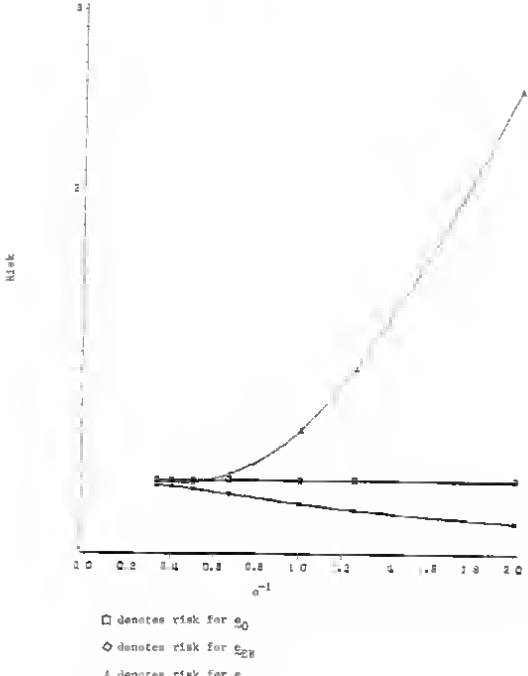


Figure 2.3 Risk Analysis for Case 3.

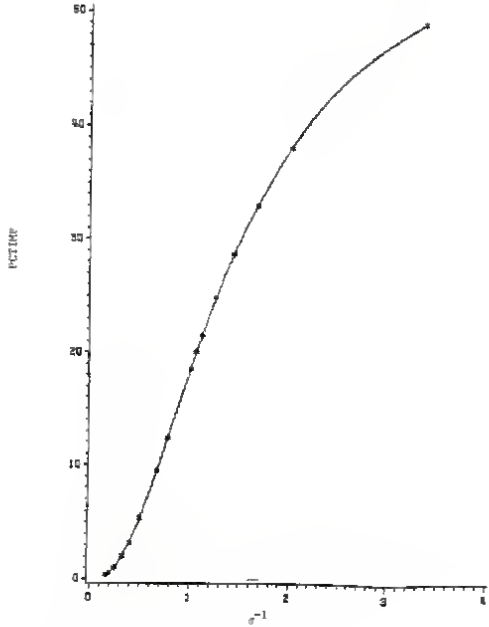


Figure 2.4 Percent Risk Improvement.

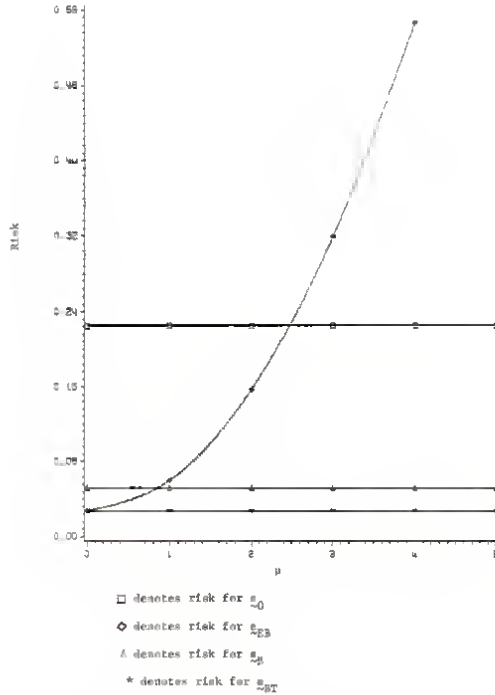


Figure 2.5 Risk Analysis for the Balanced Case

CHAPTER THREE
ESTIMATION OF MEANS FROM STRATIFIED SAMPLES IN THE PRESENCE OF AUXILIARY INFORMATION

3.1 Introduction

Auxiliary variables play an important role in survey sampling. One can typically improve on one's estimators or predictors by utilizing relevant auxiliary information. Sigman et al. (1978), Battese and Fuller (1981), Fuller and Harter (1985), Fay and Herriot (1979) among others considered relevant auxiliary information in small area estimation problems. For example, in estimating the acreage of corn and the acreage of soybean for 12 counties in North Central Iowa in 1978, Fuller and Harter (1985) used auxiliary information from the Landsat satellite in addition to the data available from U.S.D.A. on the acreage of these crops for 36 sampling units (segments) in this area. The Landsat satellite information was quantified as the number of pixels (the unit of satellite observation) classified as corn and soybean for this area. In estimating 1969 per capita income for small places, Fay and Herriot (1979) demonstrated how one can improve on one's estimators by incorporating some relevant auxiliary information in the estimation procedure. In particular, they considered county values, tax return data for 1969, and data on housing from the 1970 census in addition to the 1970 census sample estimates of per

capita income for these small places. However, the methods employed by these authors rely heavily on the normality assumption.

Suppose, in addition to the variable of interest X, we have information on an auxiliary variable, say Z_i, for all the units in each stratum. Let Z_{ji} denote the value of the auxiliary variable associated with the ith unit in the jth stratum. We assume that the Z_{ji}'s are known and $Z_{ji} > 0$ for all $i = 1,\ldots,N_j$; $j = 1,\ldots,m$. Let $Y_{ji} = (X_{ji}-c_{ji})/Z_{ji}$ for $i = 1,\ldots,N_j$; $j = 1,\ldots,m$. A model exactly similar to the one assumed for X_{ji}'s in Section 2.1 is assumed for Y_{ji}'s. In this chapter, we consider the estimation of the vector of stratum means $\underline{\gamma} = (\gamma_1,\ldots,\gamma_m)$, where $\gamma_j = N_j^{-1}\sum_{i=1}^{N_j} X_{ji}$. However, unlike the last chapter, we will take into account the auxiliary variables Z_{ji}'s in arriving at an estimator of $\underline{\gamma}$ in addition to the X_{ji}'s. We consider the loss function

$$L(\underline{a},\underline{\gamma}) = m^{-1}\sum_{j=1}^{m}(a_j-\gamma_j)^2. \qquad (3.1.1)$$

In Section 3.2, we find the Bayes estimator. Different empirical Bayes estimators are obtained in Section 3.3. These empirical Bayes estimators are compared with a generalized regression estimator in Section 3.4. We find exactly the same MSE behavior of the proposed empirical Bayes estimators found in Section 2.4 with respect to the generalized regression estimator.

3.2 Bayes Estimation

Under the loss given in (3.1.1), Bayes estimator of $\underline{\gamma}$ is $\delta_{RB} = (\delta_{RB}^{(1)},\ldots,\delta_{RB}^{(m)})$, where

$$\begin{aligned}\delta_{RB}^{(j)} &= E[Y_j|\underline{x}_j] \\ &= N_j^{-1}[\Sigma_{i=1}^{n_j} X_{ji} + \Sigma_{i=n_j+1}^{N_j} E(X_{ji}|\underline{x}_j)] \\ &= N_j^{-1}[\Sigma_{i=1}^{n_j} X_{ji} + \Sigma_{i=n_j+1}^{N_j} E\{c_{ji}+Z_{ji}E(Y_{ji}|\underline{x}_j)\}] \\ &= N_j^{-1}[\Sigma_{i=1}^{n_j} X_{ji} + \Sigma_{i=n_j+1}^{N_j}\{B_j\mu+(1-B_j)\bar{Y}_j\} + \Sigma_{i=n_j+1}^{N_j} Z_{ji}],\end{aligned} \tag{3.2.1}$$

where

$$\bar{Y}_j = o_j^{-1}\Sigma_{i=1}^{n_j} Y_{ji}.$$

Here, B_j's and μ have the same expressions as given in Section 2.2 except that now the parameters refer to the Y_{ji}'s instead of the X_{ji}'s.

3.3 Empirical Bayes Estimation

As in Section 2.3, we replace the parameters of the prior distribution in (3.2.1) by their estimators to get different empirical Bayes estimators. Define $\bar{\bar{Y}} = \Sigma_{j=1}^{m} n_j \bar{Y}_j / n_T$. We replace X's by Y's in the definitions of SSB and SSW given in Section 2.3. The definitions of MSB, MSW, $\hat{M}$, $\hat{B}_j$, $\hat{\bar{B}}$, $\hat{\hat{B}}$, μ_k and $\hat{\mu}$ would change accordingly.

Now, from (3.2.1), an empirical Bayes estimator of $\underset{\sim}{\gamma}$ is $\hat{\underset{\sim}{e}}_{REB} = (e^{(1)}_{REB},\ldots,e^{(m)}_{REB})$, where

$$e^{(j)}_{REB} = N_j^{-1}[\Sigma_{i=1}^{n_j} X_{ji} + \Sigma_{i=n_j+1}^{N_j} c_{ji} + (\hat{B}_j\hat{\mu} + (1-\hat{B}_j)\bar{Y}_j)\Sigma_{i=n_j+1}^{N_j} Z_{ji}]. \quad (3.3.1)$$

In the case when M and accordingly B_j's are known, but μ is unknown, one estimates μ by μ_* instead of $\hat{\mu}$, and the resulting empirical Bayes estimator of γ is $\hat{\underset{\sim}{e}}^{0}_{REB} = (e^{0(1)}_{REB},\ldots,e^{0(m)}_{REB})$, where

$$e^{0(j)}_{REB} = N_j^{-1}[\Sigma_{i=1}^{n_j} X_{ji} + \Sigma_{i=n_j+1}^{N_j} c_{ji} + (\hat{B}_j\mu_* + (1-\hat{B}_j)\bar{Y}_j)\Sigma_{i=n_j+1}^{N_j} Z_{ji}]. \quad (3.3.2)$$

In the balanced situation when $n_1 = \ldots = n_m = n$, if one estimates B by $\hat{B}$, and μ by $\bar{\bar{Y}}$, one gets the corresponding empirical Bayes estimator $\hat{\underset{\sim}{e}}^{*}_{REB} = (e^{*(1)}_{REB},\ldots,e^{*(m)}_{REB})$, where

$$e^{*(j)}_{REB} = N_j^{-1}[\Sigma_{i=1}^{n} X_{ji} + \Sigma_{i=n+1}^{N_j} c_{ji} + (\hat{B}\bar{\bar{Y}} + (1-\hat{B})\bar{Y}_j)\Sigma_{i=n+1}^{N_j} Z_{ji}]. \quad (3.3.3)$$

In this case, if instead one estimates B by $\hat{\hat{B}}$, then the corresponding empirical Bayes estimator is $\hat{\underset{\sim}{e}}_{RJS} = (e^{(1)}_{RJS},\ldots,e^{(m)}_{RJS})$, where

$$e^{(j)}_{RJS} = N_j^{-1}[\Sigma_{i=1}^{n} X_{ji} + \Sigma_{i=n+1}^{N_j} c_{ji} + (\hat{\hat{B}}\bar{\bar{Y}} + (1-\hat{\hat{B}})\bar{Y}_j)\Sigma_{i=n+1}^{N_j} Z_{ji}]. \quad (3.3.4)$$

Also, note that if $n_1 = \ldots = n_m = n$, and M and accordingly B is known, $\underline{\bar{Y}}$ is automatad by $\underline{\bar{e}}_{REB}^{**} = (\bar{e}_{REB}^{**(1)}, \ldots, \bar{e}_{REB}^{**(m)})$, where

$$\bar{e}_{REB}^{**(j)} = N_j^{-1}[\Sigma_{i=1}^{n} X_{ji} + \Sigma_{i=n+1}^{N_j} c_{ji} + \{\bar{n}\bar{Y} + (1-B)\bar{\Sigma}_j\}\Sigma_{i=n+1}^{N_j} Z_{ji}]. \tag{3.3.5}$$

3.4 RSL Behavior

The usual ratio estimator of $\underline{\bar{Y}}$ is $\underline{\bar{e}}_R = (\bar{e}_R^{(1)}, \ldots, \bar{e}_R^{(m)})$, where

$$\bar{e}_R^{(j)} = N_j^{-1}\{\Sigma_{i=1}^{n_j} X_{ji} / \Sigma_{i=1}^{n_j} Z_{ji}\}\Sigma_{i=1}^{N_j} Z_{ji}. \tag{3.4.1}$$

A more general class of regression estimators is given by $\underline{\bar{e}}_0 = (\bar{e}_0^{(1)}, \ldots, \bar{e}_0^{(m)})$, where

$$\bar{e}_0^{(j)} = N_j^{-1}[\Sigma_{i=1}^{n_j} X_{ji} + \Sigma_{i=n_j+1}^{N_j} c_{ji} + \Sigma_{i=1}^{n_j}\{(X_{ji} - c_{ji})/Z_{ji}\}\{d_{ji}/\Sigma_{i=1}^{n_j} d_{ji}\}\Sigma_{i=n_j+1}^{N_j} Z_{ji}]. \tag{3.4.2}$$

The above estimator can be motivated from a superpopulation approach as in Basu (1971), Royall and Herson (1973), or Scott, Brewer and Ho (1978). Such estimators have been developed using a "stepwise Bayes" approach by Ghosh and Meeden (1986) and Verdooren and Meeden (1983). Godambe (1966) have an interesting empirical Bayes interpretation of the classical ratio estimator, i.e. when $c_{ji} = 0$ and $d_{ji} = Z_{ji}$. In a special case, when $d_{ji} = d_j$ for all

$i = 1,\ldots,n_j$, $j=1,\ldots,m$. (3.4.2) leads to

$$e_{RO}^{(j)} = N_j^{-1}[\Sigma_{i=1}^{n_j} X_{ji} + \Sigma_{i=n_j+1}^{N_j} c_{ji} + \bar{Y}_j(\Sigma_{i=n_j+1}^{N_j} Z_{ji})]. \tag{3.4.3}$$

Let $\underset{\sim}{e}_{RO} = (e_{RO}^{(1)},\ldots,e_{RO}^{(m)})$.

We compare the empirical Bayes estimator proposed in Section 3.3 with $\underset{\sim}{e}_{RO}$ on the basis of RSL. The following theorem gives an expression for $RSL(\xi,\underset{\sim}{e}_{REB},\underset{\sim}{e}_{RO})$.

Theorem 3.4.1 Under the assumed model

$$RSL(\xi,\underset{\sim}{e}_{REB},\underset{\sim}{e}_{RO}) = \frac{\Sigma_{j=1}^{m} B_j^2 E[(\hat{B}_j - B_j)(\bar{Y}_j - \mu) - \hat{B}_j(\hat{\mu}-\mu)]^2}{\Sigma_{j=1}^{m} B_j^2 E(\bar{Y}_j - \mu)^2}, \tag{3.4.4}$$

where

$$Z_j^2 = N_j^{-2} \sum_{s_j \in S_j} p(s_j)\Big(\sum_{i \notin s_j} Z_{ji}\Big)^2.$$

Proof: Using (3-2.1), (3.4.3), and Lemma 2.4.1,

$$r(\xi,\underset{\sim}{e}_{RO}) - r(\xi,\underset{\sim}{e}_{REB})$$

$$= m^{-1}\Sigma_{j=1}^{m} \sum_{s_1 \in S_1} \cdots \sum_{s_m \in S_m} [\textstyle\prod_{j=1}^{m} p(s_j)] E(e_{RO}^{(j)} - e_{REB}^{(j)})^2$$

$$= m^{-1}\Sigma_{j=1}^{m} \sum_{s_1 \in S_1} \cdots \sum_{s_m \in S_m} [\textstyle\prod_{j=1}^{m} p(s_j)] N_j^{-2} B_j^2 \Big(\sum_{i \notin s_j} Z_{ji}\Big)^2 E(\bar{Y}_j - \mu)^2$$

$$= m^{-1}\Sigma_{j=1}^{m} B_j^2 E(\bar{Y}_j - \mu)^2. \tag{3.4.5}$$

Using (3.2.1), (3.3.1), and Lemma 2.4.1,

$$r(\xi,\underline{s}_{REG})-r(\xi,\underline{s}_{RH})$$

$$= m^{-1}\sum_{j=1}^{m} \sum_{s_1\in S_1} \cdots \sum_{s_m\in S_m} \{\prod_{j=1}^{m} p(s_j)N_j\}^{-2}\Big(\sum_{1\neq s_j} Z_{ji}\Big)^2$$
$$E[(\hat{B}_j-B_j)(\bar{Y}_j-\mu)+\hat{B}_j(\hat{\mu}-\mu)]^2$$

$$= m^{-1}\sum_{j=1}^{m} g_j^2 E[(\hat{B}_j-B_j)(\bar{Y}_j-\mu)+\hat{B}_j(\hat{\mu}-\mu)]^2. \tag{3.4.6}$$

Combining (3.4.5) and (3.4.6), one gets the theorem.

Now, if one assumes

$$0 < a \leq \inf_{1\leq j\leq m}\ \inf_{1\leq i\leq N_j} Z_{ji} \leq \sup_{1\leq j\leq m}\ \sup_{1\leq i\leq N_j} Z_{ji} \leq b < \infty,$$

using the fact that $\sum_{s_j\in S_j} p(s_j) = 1$, one gets

$$B_j^2 \leq b^2 f_j^2 \tag{3.4.7}$$

and

$$g_j^2 \geq a^2 f_j^2. \tag{3.4.8}$$

Using (3.4.7), (3.4.8), (3.4.4), and (2.4.4), one gets

$$RSL(\xi,\underline{s}_{REG},\underline{s}_{RO})$$

$$\leq \frac{b^2}{a^2} RSL(\xi,\underline{s}_{EG},\underline{s}_{O}). \tag{3.4.9}$$

Now, using Theorem 2.4.2, one gets the following result.

$$RSL(\xi, \underset{\sim}{e}_{REB}, \underset{\sim}{e}_{RO}) \to 0 \text{ as } m \to \infty .$$

By an argument similar to (3.4.9), one gets

$$RSL(\xi; \underset{\sim}{e}^{0}_{REB}, \underset{\sim}{e}_{RO}) \leq \frac{k^2}{n} RSL(\xi, \underset{\sim}{e}^{0}_{EB}, \underset{\sim}{e}_{O}). \tag{3.4.10}$$

Now, using (2.4.33), one gets

$$RSL(\xi; \underset{\sim}{e}^{0}_{REB}, \underset{\sim}{e}_{RO}) = O(m^{-1}) \tag{3.4.11}$$

Also, in the balanced case when $n_1 = \ldots = n_m = n$, by an argument similar to (2.4.34), one gets

$$RSL(\xi, \underset{\sim}{e}^{**}_{REB}, \underset{\sim}{e}_{RO}) = m^{-1}. \tag{3.4.12}$$

Now, if one assumes that the random vector variables $\underset{\sim}{Y}_1, \ldots, \underset{\sim}{Y}_m$ are iid and also

$$E[(SSW)^2/SSB] = E(SSW)^2/E(SSB) + O(1), \tag{3.4.13}$$

one gets

$$RSL(\xi; \underset{\sim}{e}_{RJS}, \underset{\sim}{e}_{RO}) = O(e^{-1}). \tag{3.4.14}$$

CHAPTER FOUR
ESTIMATION OF VARIANCES FROM STRATIFIED SAMPLES

4.1 Introduction

In this chapter, we consider simultaneous estimation of variances from several strata, where each stratum contains a finite number of elements. The present chapter retains the spirit of Chapter Two and considers empirical Bayes estimation when no distributional assumption is made for the superpopulation model.

We use the same notations as in Chapter Two. The problem is estimation of $\underline{\gamma} = (\gamma_1,\ldots,\gamma_m)$, where $\gamma_j=(N_j-1)^{-1}\sum_{i=1}^{N_j}(X_{ji}-\bar{X}_{N_j})^2$ ($\bar{X}_{N_j} = \sum_{i=1}^{N_j} X_{ji}$) under the loss

$$L(\underline{a},\underline{\gamma}) = m^{-1}\sum_{j=1}^{m}(a_j-\gamma_j)^2. \qquad (4.1.1)$$

In addition to the model assumed in Section 2.1, we assume

$$\propto\ V(X_{ji}|\theta_j) = c_0 + c_1\theta_j + c_2\theta_j^2 = \mu_2(\theta_j) \quad (say), \qquad (4.1.2)$$

where at least one of the c_i's is non-zero ($i=0,1,2$). This is, for example, satisfied when X_{ji}'s come from a natural exponential family with quadratic variance function (cf Morris (1983a)). As a special case, we could have $c_1 = c_2 = 0$ and $c_0 = \tau^2$ (>0), which hold, for example, in the normal superpopulation model assumed in Ericson (1969a), or in Ghosh and Meeden (1986). As a second

example, assume that conditional on θ_j, X_{ji}'s are iid Poisson (θ_j). Then $c_0 = c_2 = 0$ and $c_1 = 1$. Morris (1983a) characterizes all possible distributions belonging to the natural exponential family with quadratic variance functions. However, our analysis can incorporate other distributions as well in the superpopulation model. As an example, it holds for multivariate t described in Section 2.1.

The outline of this chapter is as follows. In Section 4.2, we have developed Bayes estimators of variances from different strata, assuming only linearity of the posterior means of the θ_j's in the sample observations. Ericson (1969b) used this idea for estimating the finite population mean, but to our knowledge, his idea has not been pursued for estimating the finite population variance. It turns out that the Bayes estimator of the finite population variance developed by Ericson (1969a) is indeed a special case of the estimator developed in Section 4.2. In Section 4.3, different empirical Bayes estimators are obtained from a classical viewpoint. The development follows the lines of Chapter Two. In Section 4.3, the risk performance of the empirical Bayes estimators is studied in the general case when sample sizes from different strata are not equal. Certain additional results are provided when sample sizes are all equal.

4.2 Bayes Estimation

Under the assumed model and the posterior linearity condition described in (2.1.3), it follows from Goldstein (1975b)

$$E(\theta_j|\underline{X}_j) = (1-B_j)\bar{X}_j + B_j\mu \tag{4.2.1}$$

$$= \mu + (1-B_j)(\bar{X}_j-\mu) \tag{4.2.2}$$

$$= \bar{X}_j - B_j(\bar{X}_j-\mu), \tag{4.2.3}$$

$$V(\theta_j|\underline{X}_j) = \tau^2 n_j^{-1}(1-B_j), \tag{4.2.4}$$

where $E[\mu_2(\theta_j)] = \tau^2$, $M = \tau^2/\sigma^2$ and $B_j = M/(M+n_j)$. Equation (4.2.1) also appears in Ericson (1969b) [see also Hartigan (1969)]. Next, writing $\gamma_j = [N_j(N_j-1)]^{-1} \sum_{1\le i\ne i'\le N_j} (X_{ji}-X_{ji'})^2/2$, it follows that under the loss (4.1.1), the Bayes estimator of $\underline{\gamma}$ is $\underline{e}_B = (e_B^{(1)},\ldots,e_B^{(m)})$, where

$$e_B^{(j)} = [N_j(N_j-1)]^{-1}E\{\Big[\sum_{1\le i\ne i'\le n_j} (X_{ji}-X_{ji'})^2/2 + 2\sum_{i=1}^{n_j}\sum_{i'=n_j+1}^{N_j}(X_{ji}-X_{ji'})^2/2 + \sum_{n_j+1\le i\ne i'\le N_j}(X_{ji}-X_{ji'})^2/2\Big]\Big|\underline{X}_j\}. \tag{4.2.5}$$

Now, for $n_j+1\le i\ne i'\le N_j$, using $\mu_2(\theta_j) = c_0 + c_1\theta_j + c_2\theta_j^2$, (4.2.2),

(4.2.4), and the fact that conditional on θ_j, X_{ji}'s are mutually independent, one gets

$$
\begin{aligned}
& E\{(X_{ji}-X_{ji'})^2 \mid \underline{X}_j\} \\
&= E\{E[(X_{ji}-X_{ji'})^2 \mid \theta_j, \underline{X}_j] \mid \underline{X}_j\} \\
&= E\{\{E[(X_{ji}-\theta_j)-(X_{ji'}-\theta_j)]^2 \mid \theta_j, \underline{X}_j\} \mid \underline{X}_j\} \\
&= 2E[\mu_2(\theta_j) \mid \underline{X}_j] \\
&= 2E[(c_0+c_1\theta_j+c_2\theta_j^2) \mid \underline{X}_j] \\
&= 2[c_0+c_1E(\theta_j \mid \underline{X}_j)+c_2\{V(\theta_j \mid \underline{X}_j)+[E(\theta_j \mid \underline{X}_j)]^2\}] \\
&= 2[c_0+c_1\{\mu+(1-B_j)(\bar{X}_j-\mu)\} \\
&\qquad +c_2[\tau^2 n_j^{-1}(1-B_j)+(1-B_j)^2(\bar{X}_j-\mu)^2+\mu^2+2\mu(1-B_j)(\bar{X}_j-\mu)\}] \\
&= 2[(c_0+c_1\mu+c_2\mu^2)+c_2\tau^2 n_j^{-1}(1-B_j) \\
&\qquad +(c_1+2c_2\mu)(1-B_j)(\bar{X}_j-\mu)+c_2(1-B_j)^2(\bar{X}_j-\mu)^2]. \qquad (4.2.6)
\end{aligned}
$$

Also, for $1 \le i \ne i' \le n_j$ and $n_j+1 \le i \ne i' \le N_j$, using (4.2.3) one gets

$$
\begin{aligned}
& E\{(X_{ji}-X_{ji'})^2 \mid \underline{X}_j\} \\
&= E\{[(X_{ji}-\theta_j) - (X_{ji'}-\theta_j)]^2 \mid \underline{X}_j\} \\
&= E[\{E[(X_{ji}-\theta_j) - (X_{ji'}-\theta_j)]^2 \mid \underline{\theta}_j, \underline{X}_j\} \mid \underline{X}_j] \\
&= E[\{(X_{ji}-\theta_j)^2 + \mu_2(\theta_j)\} \mid \underline{X}_j]
\end{aligned}
$$

$$= V(\theta_j|\underline{X}_j) + [E(\theta_j|\underline{X}_j)-X_{ji}]^2 + E(\mu_2(\theta_j)|\underline{X}_j]$$

$$= V(\theta_j|\underline{X}_j)+[B_j(\bar{X}_j-\mu)+(X_{ji}-\bar{X}_j)]^2+E[\mu_2(\theta_j)|\underline{X}_j]$$

$$= V(\theta_j|\underline{X}_j)+B_j^2(\bar{X}_j-\mu)^2+(X_{ji}-\bar{X}_j)^2+2B_j(\bar{X}_j-\mu)(X_{ji}-\bar{X}_j)$$

$$+E[\mu_2(\theta_j)|\underline{X}_j]. \qquad (4.2.7)$$

Writing

$$s_j^2 = (n_j-1)^{-1}\sum_{i=1}^{n_j}(X_{ji}-\bar{X}_j)^2$$

$$= (n_j(n_j-1))^{-1}\sum_{1\le i\ne i'\le n_j}(X_{ji}-X_{ji'})^2/2,$$

and using (4.2.6), (4.2.7), it follows from (4.2.5) that

$$e_s^{(j)} = (N_j(N_j-1))^{-1}[n_j(n_j-1)s_j^2 + (n_j-1)(N_j-n_j)s_j^2$$

$$+n_j(N_j-n_j)\{V(\theta_j|\underline{X}_j)+B_j^2(\bar{X}_j-\mu)^2\}$$

$$+(N_j-n_j)(N_j-1)E(\mu_2(\theta_j)|\underline{X}_j)]$$

$$= (N_j(N_j-1))^{-1}[N_j(n_j-1)s_j^2$$

$$+(N_j-n_j)\{(N_j-1)(c_0+c_1\mu+c_2\mu^2)$$

$$+(n_j+c_2(N_j-1))\tau^2 n_j^{-1}(1-B_j)\}$$

$$+(N_j-n_j)(N_j-1)(c_1+2c_2\mu)(1-B_j)(\bar{X}_j-\mu)$$

$$+(N_j-n_j)\{n_jB_j^2 + (N_j-1)(1-B_j)^2c_2\}(\bar{X}_j-\mu)^2]. \qquad (4.2.8)$$

Now, writing $g_j = n_j/(N_j-1)$ and $f_j = (N_j-n_j)/N_j$ (the finite population correction factor), it follows from (4.2.8) that

$$e_B^{(j)} = \{(n_j-1)/(N_j-1)\}s_j^2 + f_j[(c_0+c_1\mu+c_2\mu^2)+(g_j+c_2)\tau^2 n_j^{-1}(1-B_j) + (c_1+2c_2\mu)(1-B_j)(\bar{X}_j-\mu)+\{g_j B_j^2+c_2(1-B_j)^2\}(\bar{X}_j-\mu)^2]. \quad (4.2.9)$$

In the special case when $c_1 = c_2 = 0$ so that $c_0 = \tau^2$, one gets from (4.2.8)

$$\begin{aligned} & e_B^{(j)} \\ &= \{(n_j-1)/(N_j-1)\}s_j^2 + f_j[\tau^2(N_j-n_j)/(N_j-1) \\ &\qquad + \{n_j/(N_j-1)\}(\bar{X}_j-\mu)^2] \\ &= (N_j-1)^{-1}[(n_j-1)s_j^2 + f_j[\tau^2(N_j-n_j) + n_j(\bar{X}_j-\mu)^2]]. \end{aligned} \quad (4.2.10)$$

This estimator was derived by Nelson (1969a) under normality of the superpopulation.

4.3 Empirical Bayes Methodology

In Section 2.3, we have discussed the problem of estimation of different prior parameters. Plugging the estimators $\hat{\mu}$, $\hat{\tau}^2$ and $\hat{B}_j$ for μ, τ^2 and B_j respectively in (4.2.9), one gets the empirical Bayes estimator $e_{EB} = (e_{EB}^{(1)},\ldots,e_{EB}^{(m)})$ of χ, where

$$e_{EE}^{(j)} = \{(n_j-1)/(N_j-1)\}s_j^2$$
$$+ f_j[(c_0+c_1\hat{\mu}+c_2\hat{\mu}^2)+(g_j+c_2)n_j^{-1}(1-\hat{B}_j)MSW\}$$
$$+ f_j(c_1+2c_2\hat{\mu})(1-\hat{B}_j)(\bar{X}_j-\hat{\mu})+f_j\{g_j\hat{B}_j^2+c_2(1-\hat{B}_j)^2\}(\bar{X}_j-\hat{\mu})^2.$$
$$(j = 1,\ldots,m). \qquad (4.3.1)$$

4.4 RSL Behaviur

In chis section, we study the Bayes risk performance of δ_{EE} with the classical estimator $\delta_0 = (e_0^{(1)},..,e_0^{(m)})$ of χ, where $e_0^{(j)} = s_j^2$, $j = 1,\ldots,m$. Rather than comparing directly the Bayes risks of δ_{EE} and δ_0, we compute RSL $(\xi, \delta_{EE}, \delta_0)$ defined in (2.4.1). Using Lemma 2-4-1, it follows from (2.4.1) that

RSL $(\xi, \delta_{EE}, \delta_0)$

$$= m^{-1}\tau^m\prod_{j=1}^m f_j^2 E[\{(c_0+c_1\hat{\mu}+c_2\hat{\mu}^2) - (c_0+c_1\mu+c_2\mu^2)\}$$
$$+[(g_j+c_2)n_j^{-1}(MSW(1-\hat{B}_j) - \tau^2(1-B_j))]$$
$$+[(c_1+2c_2\hat{\mu})(1-\hat{B}_j)(\bar{X}_j-\hat{\mu}) - (c_1+2c_2\mu)(1-B_j)(\bar{X}_j-\mu)]$$
$$+[\{(g_j\hat{B}_j+c_2(1-\hat{B}_j)^2\}(\bar{X}_j-\hat{\mu})^2 - \{g_jB_j^2+c_2(1-B_j)^2\}(\bar{X}_j-\mu)^2]\}^2$$

$$+ m^{-1}\tau^m\prod_{j=1}^m f_j^2 E[-N_j(N_j-1)^{-1}s_j^2 + (c_0+c_1\mu+c_2\mu^2) + (g_j+c_2)\tau^2 e_j^{-1}(1-B_j)$$
$$+(c_1+2c_2\mu)(1-B_j)(\bar{X}_j-\mu) + \{g_jB_j^2+c_2(1-B_j)^2\}(\bar{X}_j-\mu)^2\}^2.$$
$$(4.4.1)$$

We next investigate the asymptotic behavior of the RSL as $m \to \infty$. Write $\mu_r(\theta_j) = E\{|X_{ji} - \theta_j|^r | \theta_j\}$. The following theorem is proved.

Theorem 4.4.1. Assume the superpopulation model given in Section 4.1. Then if

(i) $\inf_j n_j \geq 2$, (ii) $\sup_j n_j = K < \infty$, (iii) $E[\mu_{4+\delta}(\theta_1)] < \infty$ for some $\delta > 0$, (iv) $E|\theta_1 - \mu|^{4+\delta} < \infty$ for some $\delta > 0$, and (v) for at least one $1 < j < \infty$, the prior distribution of θ_j is not degenerate at less than three points,

$$\text{RSL}(\xi, \xi_{EB}, \xi_O) \to 0 \text{ as } m \to \infty.$$

The proof of the theorem is a consequence of the following two lemmas. We define A_m^2 and D_m^2 for the random variables given respectively in the numerator and denominator in the right hand side of (4.4.1). The following two lemmas are proved.

Lemma 4.4.1 Under the assumptions (i), (ii), and (v) of Theorem 4.4.1 along with

(iii) $0 < E[\mu_2(\theta_1)] < \infty$ and (iv) $0 < \sigma^2 = V(\theta_1) < \infty$,

$$\liminf_{m \to \infty} E(D_m^2) > 0.$$

Lemma 4.4.2 Under the assumptions of Theorem 4.4.1,

$$E(A_m^2) \to 0 \text{ as } m \to \infty$$

Note that the conditions (iii) and (iv) of Lemma 4.4.1 are

milder than (iii) and (iv) of Theorem 4.4.1. Also, condition (v) of Theorem 4.4.1 is very mild--it holds, for example, for any continuous distribution. It follows from Lemma 4.2.2 that the proposed empirical Bayes estimator is asymptotically optimal in the sense of Robbins (1955).

Proof of Lemma 4.4.1

First use the inequality

$$E(D_m^2) \geq V(D_m) \geq V(E(D_m|\underline{\theta}))$$

$$= m^{-1}\sum_{j=1}^{m} f_j^2\, V\{-N_j(N_j-1)^{-1}\mu_2(\theta_j) + (\theta_j-\mu)(c_1+2c_2\mu)(1-\theta_j) + [c_2\theta_j^2+c_2(1-\theta_j)^2][(\theta_j-\mu)^2+n_j^{-1}\mu_2(\theta_j)]\}. \tag{4.4.2}$$

Note that $\mu_2(\theta_j) = c_0 + c_1\theta_j + c_2\theta_j^2$. Write

$$d_{j1} = -c_1N_j(N_j-1)^{-1}+(c_1+2c_2\mu)(1-\theta_j)+[\theta_j\theta_j^2+c_2(1-\theta_j)^2](-2\mu+c_1n_j^{-1}),$$
$$d_{j2} = -c_2N_j(N_j-1)^{-1}+[\theta_j\theta_j^2+c_2(1-\theta_j)^2](1+c_2n_j^{-1}). \tag{4.4.3}$$

Then, it follows from (4.4.2) and (4.4.3) that

$$E(D_m^2) \geq m^{-1}\sum_{j=1}^{m} f_j^2 V(d_{j1}\theta_j+d_{j2}\theta_j^2). \tag{4.4.4}$$

The lemma follows immediately from condition (v) of Theorem 4.4.1.

Proof of Lemma 4.4.2 We first use the c_s -inequality, and the fact that $f_j \leq 1$ for all j to get

$$E(A_m^2)$$

$$\leq 4E\{[c_1(\hat{\mu}-\mu)+c_2(\hat{\mu}^2-\mu^2)]^2+m^{-1}\Sigma_{j=1}^m(\hat{B}_j+c_2)^2n_j^{-2}(MSW(1-\hat{B}_j)-\tau^2(1-B_j))^2$$

$$+m^{-1}\Sigma_{j=1}^m\{(c_1+2c_2\hat{\mu})(1-\hat{B}_j)(\bar{X}_j-\hat{\mu})-(c_1+2c_2\mu)(1-B_j)(\bar{X}_j-\mu)\}^2$$

$$+m^{-1}\Sigma_{j=1}^m\{[B_j\hat{B}_j^2+c_2(1-\hat{B}_j)^2](\bar{X}_j-\hat{\mu})^2-[B_jB_j^2+c_2(1-B_j)^2](\bar{X}_j-\mu)^2\}^2\}. \quad (4.4.5)$$

To prove that the right hand side of (4.4.5) converges to zero, first use the following inequalities

(i) $(\hat{\mu}-\mu)^2 \leq (K^2/4)m^{-1}\Sigma_{j=1}^m(\bar{X}_j-\mu)^2$ (cf (A.14) Ghosh and Meeden (1986)), (4.4.6)

(ii) $(\hat{\mu}^2-\mu^2)^2=[(\hat{\mu}-\mu)^2+2\mu(\hat{\mu}-\mu)]^2 \leq 2\{(\hat{\mu}-\mu)^4+4\mu^2(\hat{\mu}-\mu)^2\}$, (4.4.7)

(iii) $(B_j+c_2)^2e_j^{-2} \leq \frac{1}{4}\{1+|c_2|\}^2$, (4.4.8)

(iv) $[MSW(1-\hat{B}_j)-\tau^2(1-B_j)]^2=[(MSW-\tau^2)(1-\hat{B}_j)+\tau^2(B_j-\hat{B}_j)]^2$

$$\leq 2\{(MSW-\tau^2)^2+\tau^4 \max_{1\leq j\leq m}(\hat{B}_j-B_j)^2\}, \quad (4.4.9)$$

(v) $[(c_1+2c_2\hat{\mu})(1-\hat{B}_j)(\bar{X}_j-\hat{\mu})-(c_1+2c_2\mu)(1-B_j)(\bar{X}_j-\mu)]^2$

$$\leq 2[c_1^2\{(1-\hat{B}_j)(\bar{X}_j-\hat{\mu})-(1-B_j)(\bar{X}_j-\mu)\}^2$$

$$+4c_2^2\{\hat{\mu}(1-\hat{B}_j)(\bar{X}_j-\hat{\mu})-\mu(1-B_j)(\bar{X}_j-\mu)\}^2]$$

$$\leq 4c_1^2\{(\hat{\mu}-\mu)^2(1-\hat{B}_j)^2+(B_j-\hat{B}_j)^2(\bar{X}_j-\mu)^2\}$$

$$+16c_2^2[(\hat{\mu}-\mu)^2(1-\hat{B}_j)^2(\bar{X}_j-\hat{\mu})^2+\mu^2\{(1-\hat{B}_j)(\bar{X}_j-\hat{\mu})-(1-B_j)(\bar{X}_j-\mu)\}^2]$$

$$\leq 4c_1^2[(\hat{\mu}-\mu)^2 + 4c_1^2\{\max_{1\leq j\leq m}(\bar{B}_j-\hat{B}_j)^2\}(\bar{X}_j-\mu)^2$$

$$+ 32c_2^2\{(\hat{\mu}-\mu)^4 + (\hat{\mu}-\mu)^2(\bar{X}_j-\mu)^2\}$$

$$+32c_2^2\mu^2\{(\hat{\mu}-\mu)^2+\max_{1\leq j\leq m}(\bar{B}_j-\hat{B}_j)^2(\bar{X}_j-\mu)^2\},$$

(vi)
$$\{(\bar{B}_j\hat{B}_j^2 + c_2(1-\hat{B}_j)^2)(\bar{X}_j-\hat{\mu})^2 - (\bar{B}_j\bar{B}_j^2 + c_2(1-\bar{B}_j)^2)(\bar{X}_j-\mu)^2\}^2$$

$$\leq 2(\bar{B}_j\hat{B}_j^2 + c_2(1-\hat{B}_j)^2)^2\{(\bar{X}_j-\hat{\mu})^2-(\bar{X}_j-\mu)^2\}^2$$

$$+ 2\{(\bar{B}_j(\bar{B}_j^2-\hat{B}_j^2) + c_2[(1-\bar{B}_j)^2-(1-\hat{B}_j)^2])^2(\bar{X}_j-\mu)^4\}$$

$$\leq 2(1+|c_2|)^2\{(\hat{\mu}-\mu)^2-2(\bar{X}_j-\mu)(\hat{\mu}-\mu)\}^2$$

$$+ 4[\bar{B}_j^2(\bar{B}_j^2-\hat{B}_j^2)^2+c_2^2\{(1-\bar{B}_j)^2-(1-\hat{B}_j)^2\}^2](\bar{X}_j-\mu)^4$$

$$\leq 4(1+|c_2|)^2\{(\hat{\mu}-\mu)^4 + 4(\bar{X}_j-\mu)^2(\hat{\mu}-\mu)^2\}$$

$$+ 4\{(\bar{B}_j^2-\hat{B}_j^2)^2 + \{(1-\bar{B}_j)^2-(1-\hat{B}_j)^2\}^2\}(\bar{X}_j-\mu)^4. \qquad (4.4.11)$$

It is proved in Section 2.4 that under assumptions weaker than those of Theorem 4.4.1, $\hat{\mu} \xrightarrow{P} \mu$, MSW $\xrightarrow{P} \tau^2$ and $\max_{1\leq j\leq m}|\hat{B}_j-\bar{B}_j| \xrightarrow{P} 0$ as $n\to\infty$. Moreover, under the assumptions $E[\psi_4(\theta_1)] < \infty$ and $E|\theta_1-\mu|^4 < \infty$, using the c_δ inequality it is clear that $m^{-1}\sum_{j=1}^m|\bar{X}_j-\mu|^r = O_p(1)$ as $n\to\infty$ for all $0\leq r\leq 4$. Utilizing the above information, it follows from (4.4.5)-(4.4.11) that the expression given within the brackets in the right hand side of (4.4.5) converges in probability to zero as $n\to\infty$. Next use the assumptions that $E[\psi_{4+\delta}(\theta_1)] < \infty$ and $E|\theta_1-\mu|^{4+\delta} < \infty$ for some

$\delta > 0$. Then $e^{-1}\sum_{j=1}^{m}\left|\bar{X}_j-\mu\right|^r$ is uniformly integrable in m for all $0 \leq r \leq 4$. Now, using the inequalities (4.4.6) - (4.4.11) and $\max_{1\leq j\leq m}\left|\hat{\mathbb{B}}_j-\mathbb{B}_j\right| \leq 1$, it follows that the right hand side of (4.4.5) converges to zero as $m\to\infty$. The proof of Lemma 4.4.2 is now complete.

Next, consider the case when $n_1 = \ldots = n_m = n$ (say). In this case, it is possible to obtain the rate at which RSL (ξ, ξ_{EB}, ξ_0) converges to zero as $m\to\infty$ under certain additional assumptions. First note that in this case $\hat{\mathbb{B}}_1 = \ldots = \hat{\mathbb{B}}_m = \hat{\mathbb{B}}$, where $\hat{\mathbb{B}} = \min(1, \hat{\hat{\mathbb{B}}})$ and $\hat{\hat{\mathbb{B}}} = (m-3)SSW/[m(n-1)SSB]$. Also, in this case $\hat{\mu} = m^{-1}\sum_{j=1}^{m}\bar{X}_j = \bar{\bar{X}}$. We now prove the following theorem.

Theorem 4.4.2. Let $n_1 = \ldots = n_m = n$. Assume the model given in Section 4.1 with the additional assumption that $X_{j1}, \ldots, X_{jn}$ are iid. Suppose that conditions (i)-(v) of Theorem 4.4.1 hold. Assume also that

$$E\{(SSW)^2/SSB\} = E(SSW)^2/E(SSB) + O(1), \tag{4.4.12}$$

$$E\{(\hat{\hat{\mathbb{B}}}-\mathbb{B})^2\, m^{-1}\textstyle\sum_{j=1}^{m}(\bar{X}_j-\bar{\bar{X}})^4\} = O(m^{-1}). \tag{4.4.13}$$

Then,

$$RSL\ (\xi, \xi_{EB}, \xi_0) = O(e^{-1}) \text{ as } m\to\infty.$$

Proof of Theorem 4.4.2. As before, write

$$RSL(\xi, \xi_{EB}, \xi_0) = E(A_m^2)/E(B_m^2).$$

Using Lemma 4.4.2, $\liminf_{m\to\infty} E(D_m^2) > 0$. Hence, it suffices to show

that when $n_1 = \cdots = n_m$, under the assumptions of Theorem 4.4.2,

$$E(A_m^2) = O(m^{-1}) \text{ as } m \to \infty. \tag{4.4.14}$$

Thus, in this case

$$\text{(i) } E(\hat{\mu}-\mu)^2 = E(\bar{\bar{X}}-\mu)^2 = m^{-1}V(\bar{X}_1) = O(m^{-1}), \tag{4.4.15}$$

$$\text{(ii) } E(\hat{\mu}^2-\mu^2)^2 = E\{(\hat{\mu}-\mu)^2 + 2\mu(\hat{\mu}-\mu)\}^2$$
$$\leq 2E\{(\hat{\mu}-\mu)^4 + 4\mu^2(\hat{\mu}-\mu)^2\} = O(m^{-1}); \tag{4.4.16}$$

$$\text{(iii) } E\{MSW(1-\hat{B}) - \tau^2(1-B)\}^2$$
$$\leq 2E\{(MSW-\tau^2)^2 + \tau^4(\hat{B}-B)^2\} = O(m^{-1}) ; \tag{4.4.17}$$

using $V(MSW) = O(m^{-1})$ and $E(\hat{B}-B)^2 = O(m^{-1})$

$$\text{(iv) } E\{(c_1+2c_2\bar{\bar{X}})(1-\hat{B})(\bar{X}_j-\bar{\bar{X}}) - (c_1+2c_2\mu)(1-B)(\bar{X}_j-\mu)\}^2$$
$$\leq 2\{c_1^2E\{(1-\hat{B})(\bar{X}_j-\bar{\bar{X}}) - (1-B)(\bar{X}_j-\mu)\}^2$$
$$+ 4c_2^2E\{\bar{\bar{X}}(1-\hat{B})(\bar{X}_j-\bar{\bar{X}}) - \mu(1-B)(\bar{X}_j-\mu)\}^2\}$$
$$\leq 4c_1^2E\{(\bar{\bar{X}}-\mu)^2 + (\hat{B}-B)^2(\bar{X}_j-\mu)^2\}$$
$$+16c_2^2E\{(\bar{\bar{X}}-\mu)^2(\bar{X}_j-\bar{\bar{X}})^2 + \mu^2\{(1-\hat{B})(\bar{X}_j-\bar{\bar{X}}) - (1-B)(\bar{X}_j-\mu)\}^2\}$$
$$\leq 4c_1^2E\{(\bar{\bar{X}}-\mu)^2 + (\hat{B}-B)^2(\bar{X}_j-\mu)^2\}$$
$$+ 16c_2^2E\{(\bar{\bar{X}}-\mu)^2(\bar{X}_j-\bar{\bar{X}})^2$$
$$+ 2\mu^2\{(\hat{B}-B)^2(\bar{X}_j-\bar{\bar{X}})^2 + (1-B)^2(\bar{\bar{X}}-\mu)^2\}\}, \tag{4.4.18}$$

Since $X_1, \ldots, X_m$ are iid and $(\tilde{B}-B)^2 \leq (\hat{\tilde{B}}-B)^2$, following the arguments similar to (2.4.38), one gets

$$\begin{aligned} & m^{-1}\Sigma_{j=1}^{m}\{\text{right hand side of (4.4.18)}\} \\ & \leq 4c_1^2\{E(\bar{\tilde{X}}-\mu)^2 + E\{(\hat{\tilde{B}}-B)^2(mn)^{-1}SSB\}\} \\ & \quad +16c_2^2E\{(\bar{\tilde{X}}-\mu)^2(mn)^{-1}SSB\} + 32c_2^2\mu^2E(\bar{\tilde{X}}-\mu)^2 \\ & \quad + 32c_2^2\mu^2E\{(mn)^{-1}(\hat{\tilde{B}}-B)^2SSB\}. \end{aligned} \tag{4.4.19}$$

Following Section 2.4 (see (2.4.46)), it follows that under the assumption (4.4.12), $E(\hat{\tilde{B}}-B)^2SSB = O(1)$. Also, by (4.4.15) $E(\bar{\tilde{X}}-\mu)^2 = O(m^{-1})$. Moreover, using the Schwarz inequality,

$$\begin{aligned} & E\{(\bar{\tilde{X}}-\mu)^2m^{-1}(SSB)\} \\ & \leq E^{1/2}(\bar{\tilde{X}}-\mu)^4\, E^{1/2}(SSB/m)^2 \leq E^{1/2}(\bar{\tilde{X}}-\mu)^4\, E^{1/2}(MSB)^2 \\ & = E^{1/2}(\bar{\tilde{X}}-\mu)^4\{V(MSB) + (E(MSB))^2\}^{1/2} \\ & = O(m^{-1})O(1) = O(m^{-1}), \end{aligned} \tag{4.4.20}$$

$$\begin{aligned} \text{(v)}\ & E[\{B_j\hat{\tilde{B}}^2 + c_2(1-\hat{\tilde{B}})^2\}(\bar{\tilde{X}}_j-\bar{\tilde{X}})^2 - \{B_jB^2 + c_2(1-B)^2\}(\bar{\tilde{X}}_j-\mu)^2\}^2 \\ & \leq 2E[\{B_j(\hat{\tilde{B}}^2-B^2) + c_2[(1-\hat{\tilde{B}})^2-(1-B)^2]\}^2(\bar{\tilde{X}}_j-\bar{\tilde{X}})^4 \\ & \qquad +\{B_jB^2 + c_2(1-B)^2\}^2\{(\bar{\tilde{X}}_j-\mu)^2 - (\bar{\tilde{X}}_j-\bar{\tilde{X}})^2\}^2\} \end{aligned}$$

$$\leq 4(1+|c_2|)^2(\hat{\bar{\theta}}-\theta)^2(\bar{\bar{X}}_j-\bar{\bar{X}})^4$$
$$+ (1+|c_2|)^2[(\bar{\bar{X}}-\mu)^2 + 2(\bar{\bar{X}}-\mu)(\bar{\bar{X}}_j-\bar{\bar{X}})]^2\}$$
$$\leq 4(1+|c_2|)^2[2(\hat{\bar{\theta}}-\theta)^2(\bar{\bar{X}}_j-\bar{\bar{X}})^4 + (\bar{\bar{X}}-\mu)^4$$
$$+ 4(\bar{\bar{X}}-\mu)^2(\bar{\bar{X}}_j-\bar{\bar{X}})^2]. \qquad (4.4.21)$$

Using the assumption (4.4.13), and the facts that $E(\bar{\bar{X}}-\mu)^4 = O(m^{-2})$, $E(\bar{\bar{X}}-\mu)^2 = O(m^{-1})$, $E(MSB)^2 = O(1)$, it follows that

$$m^{-1}\Sigma_{j=1}^{m} f_j^2 \text{ (right hand side of (4.4.21))} = O(m^{-1}).$$

The proof of Theorem 4.4.2 now follows from (4.4.15), (4.4.16), (4.4.17), (4.4.20) and (4.4.21).

Note that condition (4.4.12) is also needed in Section 2.4 for proving corresponding BSL result while estimating simultaneously means of several strata. Condition (4.4.13) seems ad hoc, but using Schwarz inequality, one gets

$$E[(\hat{\bar{\theta}}-\theta)^2 m^{-1}\Sigma_{j=1}^{m}(\bar{\bar{X}}_j-\bar{\bar{X}})^4]$$
$$\leq E^{1/2}(\hat{\bar{\theta}}-\theta)^4 \; E^{1/2}[m^{-1}\Sigma_{j=1}^{m}(\bar{\bar{X}}_j-\bar{\bar{X}})^4]^2$$
$$\leq E^{1/2}(\hat{\bar{\theta}}-\theta)^4 \; E^{1/2}[m^{-1}\Sigma_{j=1}^{m}(\bar{\bar{X}}_j-\bar{\bar{X}})^8]$$
$$\leq E^{1/2}(\hat{\bar{\theta}}-\theta)^4 \; E^{1/2}[m^{-1}\Sigma_{j=1}^{m}(\bar{\bar{X}}_j-\mu)^8 + (\bar{\bar{X}}-\mu)^8]. \qquad (4.4.22)$$

Thus, using (4.4.22), one can verify the condition (4.4.13) if one assumes $d = 4$ in the conditions (iii) and (iv) of Theorem 4.4.1 and also

$$E(\hat{\hat{B}}-B)^4 = O(m^{-2}). \tag{4.4.23}$$

Equation (4.4.25) holds, for example, when the superpopulation is assumed to be multivariate normal as the following computations show:

Recalling that $\hat{\hat{B}} = c(SSW/SSB)$, where $c = (n-3)/m(n-1)$, and using the independence of SSW and SSB and also the facts that $SSW \sim \tau^2\chi^2_{m(n-1)}$ and $SSB \sim B^{-1}\tau^2\chi^2_{m-1}$, one gets

$$\begin{aligned}
&E(\hat{\hat{B}}-B)^4 \\
&= E(\hat{\hat{B}}^4)-4BE(\hat{\hat{B}}^3)+6B^2E(\hat{\hat{B}}^2)-4B^3E(\hat{\hat{B}})+B^4 \\
&= c^4E(SSW)^4E(SSB)^{-4}-4Bc^3E(SSW)^3E(SSB)^{-3} \\
&\quad +6B^2c^2E(SSW)^2E(SSB)^{-2}-4cB^3E(SSW)E(SSB)^{-1}+B^4 \\
&= B^4[c^4E(\chi^2_{m(n-1)})^4E(\chi^2_{m-1})^{-4}-4c^3E(\chi^2_{m(n-1)})^3E(\chi^2_{m-1})^{-3} \\
&\quad +6c^2E(\chi^2_{m(n-1)})^2E(\chi^2_{m-1})^{-2}-4cE(\chi^2_{m(n-1)})E(\chi^2_{m-1})^{-1}+1].
\end{aligned} \tag{4.4.24}$$

Now, using the fact that

$$E(\chi^2_{2d})^r = 2^r\frac{\Gamma(d+r)}{\Gamma(d)}$$

one gets from (4.4.24)

$$
\begin{aligned}
&E(\hat{\beta}-\beta)^4/\beta^4 \\
&= \frac{(m-3)^3[m(n-1)+6][m(n-1)+4][m(n-1)+2]}{m^3(n-1)^3(m-5)(m-7)(m-9)} \\
&\quad - 4\,\frac{(m-3)^2[m(n-1)+4][m(n-1)+2]}{m^2(n-1)^2(m-5)(m-7)} \\
&\quad + 6\,\frac{(m-3)[m(n-1)+2]}{m(n-1)(m-5)} - 3 \\
&= [(m-3)^3[m(n-1)+6][m(n-1)+4][m(n-1)+2] \\
&\quad -4m(n-1)[m(n-1)+2][m(n-1)+4](m-3)^2(m-9) \\
&\quad +6m^2(n-1)^2[m(n-1)+2](m-3)(m-7)(m-9) \\
&\quad -3m^3(n-1)^3(m-5)(m-7)(m-9)] \\
&\quad + m^3(n-1)^3(m-5)(m-7)(m-9). \qquad (4.4.25)
\end{aligned}
$$

Now, we have verified that the coefficients of n^6 and n^5 in (4.4.25) vanish. Hence, the condition (4.4.23) holds when the superpopulation is multivariate normal. Following a similar argument, one can verify (4.4.23) for a multivariate t prior.

CHAPTER FIVE
ESTIMATION OF MEANS FROM TWO-STAGE
STRATIFIED SAMPLES

5.1 Introduction

In the previous chapters, we have considered estimation in finite population sampling when samples are drawn in one stage. In this chapter, we generalise the idea by estimating the means of several strata when samples are drawn from each stratum in two stages. Suppose that each unit within a stratum (called a primary unit) can be divided into a number of smaller units or subunits. A sample of primary units is selected at the first stage. If the subunit within a selected primary unit give similar results, it seems uneconomical to measure them all. A common practice is then to select and measure a sample of subunits from the chosen primary unit. The technique is called subsampling or two-stage sampling. A very common example is household surveys where counties within a state are sometimes used as the different strata. A sample of blocks within each of the counties is drawn at the first stage, and a sample of dwellings from within the chosen blocks is drawn at the second stage.

Scott and Smith (1969) carried out a Bayesian analysis for two-stage sampling with a single stratum, assuming a normal prior. Their results were generalised to three stage sampling by Malec and Sedransk (1985). Like the previous chapters, the

present chapter dispenses with the normality assumption of the prior, and assumes instead "posterior linearity" of the strata means, i.e. the posterior expectation of every stratum mean is a linear function of the sample observations.

Bayes estimators of strata means in two-stage sampling are derived in Section 5.2 under the assumption of posterior linearity. The estimators are compared with the ones of Scott and Smith (1969) and of Ericson (1976). The latter, like ours, relaxes the normality assumption of the prior, but proceeds with different initial assumptions. Different empirical Bayes estimators are obtained in Section 5.3 by replacing one or more parameters appearing in the Bayes estimator of Section 5.2 by their corresponding counterpart. Properties of the empirical Bayes estimators are discussed in Section 5.4, and certain asymptotic results are proved.

Suppose that there are K strata, where the kth stratum contains M_k primary units. Suppose also that the jth primary unit within the kth stratum contains N_{kj} secondary units or subunits. Let $N_k = \sum_{j=1}^{M_k} N_{kj}$. Let X_{kji} denote the value of the characteristic of interest for the ith subunit within the jth primary unit from the kth stratum ($i=1,\ldots,N_{kj}$, $j=1,\ldots,M_k$, $k=1,\ldots,K$). It is assumed that the X_{kji}'s are independent for different k. The following prior distribution is assumed for the X_{kji}'s.

A. Conditional on θ_{kj}'s, X_{kji}'s are iid with distribution depending only on θ_{kj}, and $E(X_{kji}|\theta_{kj}) = \theta_{kj}$, and $V(X_{kji}|\theta_{kj}) = \mu_2(\theta_{kj})$, $(i=1,\ldots,X_{kj}, j=1,\ldots,M_k)$.

B. Conditional on μ_k's, θ_{kj}'s are iid with distribution depending only on μ_k, and $E(\theta_{kj}|\mu_k) = \mu_k$, and $V(\theta_{kj}|\mu_k) = m_2(\mu_k)$.

C. The μ_k's are iid with $E(\mu_k) = \nu$ and $V(\mu_k) = \delta^2$. Also, let $\tau^2 = E\mu_2(\theta_{kj})$ and $\sigma^2 = Em_2(\mu_k)$. (5.1.1)

As an example where the assumptions (A)-(C) hold, consider the nested one way variance components model given by

$$X_{kji} = \nu + \xi_k + \eta_{kj} + \varepsilon_{kji}, \tag{5.1.2}$$

where ξ_k's, η_{kj}'s, and ε_{kji}'s are all independently distributed. Also, ε_{kji}'s are iid with $E(\varepsilon_{kji}) = 0$ and $V(\varepsilon_{kji}) = \tau^2$, η_{kj}'s are iid with $E(\eta_{kj}) = 0$ and $V(\eta_{kj}) = \sigma^2$. ξ_k's are iid with $E(\xi_k) = 0$ and $V(\xi_k) = \delta^2$. Then, writing $\mu_k = \nu + \xi_k$ and $\theta_{kj} = \nu + \xi_k + \eta_{kj} = \mu_k + \eta_{kj}$, it is easy to see that the conditions (A)-(C) hold with $\mu_2(\theta_{kj}) = \tau^2$ for all j and k, and $m_2(\mu_k) = \sigma^2$ for all k.

Our target in this chapter is estimation of $\underline{\gamma} = (\gamma_1,\ldots,\gamma_K)$, where $\gamma_k = \bar{X}_k = M_k^{-1}\Sigma_{j=1}^{M_k} X_{kj}$ X_{kji} under the loss

$$L(\underline{a},\underline{\gamma}) = K^{-1}\Sigma_{k=1}^{K}(\gamma_k - a_k)^2. \tag{5.1.3}$$

For the kth stratum, a sample of m_k primary units is taken. For the jth primary unit within the kth stratum, a sample of n_{kj}

secondary units are taken. Without loss of generality, we denote the sample values by X_{kji} $(i=1,\ldots,n_{kj},\ j=1,\ldots,m_k,\ k=1,\ldots,K)$.

5.2 Bayes Estimates

Like the previous chapters, the key assumption is that of posterior linearity, i.e writing

$$\underset{\sim}{X}_k = (X_{k11},\ldots,X_{k1n_{k1}},\ldots,X_{km_k1},\ldots,X_{km_kn_{km_k}})',$$

we assume that

$$E(\theta_{kj}|\underset{\sim}{X}_k) = \sum_{l=1}^{m_k}\sum_{i=1}^{n_{kl}} a_{kli(*)}^{(j)} X_{kli} + a_{k0(*)}^{(j)}, \qquad (5.2.1)$$

$$j=1,\ldots,M_k,\ k=1,\ldots,K.$$

The coefficients $a_{kli(*)}^{(j)}$'s and $a_{k0(*)}^{(j)}$ are obtained by minimizing

$$\sum_{j=1}^{M_k} E(\theta_{kj} - \sum_{l=1}^{m_k}\sum_{i=1}^{e_{kl}} a_{kli}^{(j)} X_{kli} - a_{k0}^{(j)})^2 \qquad (5.2.2)$$

wrt $a_{kli}^{(j)}$'s and $a_{k0}^{(j)}$'s. The minimization can be achieved by using the principle of least squares, and actually carried out in a similar context by Goldstein (1975a). Since the minimization has to be done individually for each one of k strata, we will find it convenient to drop the suffix k in (5.2-2), and try to minimize

$$\sum_{j=1}^{M} E(\theta_j - \sum_{l=1}^{m}\sum_{i=1}^{n_l} a_{li}^{(j)} X_{li} - a_0^{(j)})^2 \qquad (5.2.3)$$

with respect to $a_{li}^{(j)}$'s and $a_0^{(j)}$. We will bring in suffix k later when we actually estimate the strata means simultaneously.

To achieve the desired minimization, we use the assumptions (A)-(C) with the suffix k dropped. First note that

$$V(\theta_1) = EV(\theta_1|\mu) + VE(\theta_1|\mu)$$

$$= Em_2(\theta_1) + V(\mu)$$

$$= \sigma^2 + \delta^2 \tag{5.2.4}$$

and

$$Cov(\theta_1,\theta_{1'}) = ECov(\theta_1,\theta_{1'}|\mu) + Cov\{E(\theta_1|\mu),E(\theta_{1'}|\mu)\}$$

$$= \delta^2. \tag{5.2.5}$$

Next, using (5.2.4), (5.2.5), and the assumptions (A)-(C), one gets

$$E(X_{1i}) = EE(E(X_{1i}|\theta_1)|\mu) = EE(\theta_1|\mu) = E(\mu) = \nu, \tag{5.2.6}$$

$$V(X_{1i}) = EV(X_{1i}|\theta_1) + VE(X_{1i}|\theta_i)$$

$$= Em_2(\theta_1) + V(\theta_1)$$

$$= \tau^2 + \sigma^2 + \delta^2. \tag{5.2.7}$$

$$Cov(X_{1i},X_{1i'}) = ECov(X_{1i},X_{1i'}|\theta_1) + Cov\{E(X_{1i}|\theta_1),E(X_{1i'}|\theta_{1'})\}$$

$$= V(\theta_1) = \sigma^2 + \delta^2, \tag{5.2.8}$$

$$\mathrm{Cov}(X_{1i},X_{1'i'})$$

$$= E\mathrm{cov}(X_{1i},X_{1'i'}|\theta_1,\theta_{1'}) + \mathrm{Cov}\{E(X_{1i}|\theta_1),E(X_{1'i'}|\theta_{1'})\}$$

$$= \mathrm{Cov}(\theta_1,\theta_{1'}) = \delta^2 \qquad (5.2.9)$$

so that

$$E(X_{1i}^2) = \sigma^2 + \tau^2 + \delta^2 + \nu^2 \ , \qquad (5.2.10)$$

$$E(X_{1i}X_{1i'}) = \tau^2 + \delta^2 + \nu^2 \quad (i\neq i') \ ; \qquad (5.2.11)$$

$$E(X_{1i}X_{1'i'}) = \delta^2 + \nu^2 \quad (1\neq 1'). \qquad (5.2.12)$$

By similar arguments, one gets

$$E(\theta_j X_{j1}) = \sigma^2 + \delta^2 + \nu^2 \ ; \qquad (5.2.13)$$

$$E(\theta_j X_{i1}) = \delta^2 + \nu^2 \quad (j\neq 1). \qquad (5.2.14)$$

Then, using Theorem 2.1 of Goldstein (1975a), it follows that for $1<j<m$,

$$E(\theta_j|\underset{\sim}{X}) = \underset{\sim}{X}'\underset{\sim}{D}_0^{-1}\underset{\sim}{b}_0 \ , \qquad (5.2.15)$$

where $\underset{\sim}{X}' = (X_{11},\ldots,X_{1n_1},\ldots,X_{m1},\ldots X_{mn_m})$,

$$\underset{\sim}{D}_0 = \begin{bmatrix} 1 & \nu \underset{\sim}{1}_{n_1}' & \cdots & \nu \underset{\sim}{1}_{n_j}' & \cdots & \nu \underset{\sim}{1}_{n_m}' \\ \nu \underset{\sim}{1}_{n_1} & \tau^2 \underset{\sim}{I}_{n_1} + (\sigma^2+\alpha^2) \underset{\sim}{1}_{n_1} \underset{\sim}{1}_{n_1}' & \cdots & \sigma^2 \underset{\sim}{1}_{n_1} \underset{\sim}{1}_{n_m}' & \cdots & \sigma^2 \underset{\sim}{1}_{n_1} \underset{\sim}{1}_{n_m}' \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \nu \underset{\sim}{1}_{n_j} & \sigma^2 \underset{\sim}{1}_{n_j} \underset{\sim}{1}_{n_1}' & \cdots & \tau^2 \underset{\sim}{I}_{n_j} + (\sigma^2+\alpha^2) \underset{\sim}{1}_{n_j} \underset{\sim}{1}_{n_j}' & \cdots & \sigma^2 \underset{\sim}{1}_{n_j} \underset{\sim}{1}_{n_m}' \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \nu \underset{\sim}{1}_{n_m} & \sigma^2 \underset{\sim}{1}_{n_m} \underset{\sim}{1}_{n_1}' & \cdots & \sigma^2 \underset{\sim}{1}_{n_m} \underset{\sim}{1}_{n_j}' & \cdots & \tau^2 \underset{\sim}{I}_{n_m} + (\sigma^2+\alpha^2) \underset{\sim}{1}_{n_m} \underset{\sim}{1}_{n_m}' \end{bmatrix} \tag{5.2.16}$$

$$\underset{\sim}{k}_0' = [\nu \quad \sigma^2 \underset{\sim}{1}_{n_1}' \ldots (\sigma^2+\alpha^2) \underset{\sim}{1}_{n_j}' \ldots \sigma^2 \underset{\sim}{1}_{n_m}'], \tag{5.2.17}$$

where $\sigma^2 = \tau^2 + \nu^2$. Throughout we use the notation $\underset{\sim}{1}_p$ for an identity matrix of order p, and $\underset{\sim}{1}_p$ for a p-component column vector with all its element equal to 1. From (5.1.15)-(5.1.17), it is easy to observe that $\underset{\sim}{k}_0' \underset{\sim}{D}_0^{-1} \underset{\sim}{k}_0$ is indeed a function of $\underset{\sim}{\bar{x}} = (\underset{\sim}{\bar{x}}_1, \ldots, \underset{\sim}{\bar{x}}_m)'$ with $\underset{\sim}{\bar{x}}_j = n_j^{-1} \underset{i=1}{\overset{n_j}{\Sigma}} \underset{\sim}{x}_{ji}$. Thus, we have

$$E(\theta_j | \underset{\sim}{x}) - E(\theta_j | \underset{\sim}{\bar{x}}) = \underset{l=1}{\overset{m}{\Sigma}} d_{1(*)}^{(j)} \underset{\sim}{\bar{x}}_l + d_{0(*)}^{(j)}, \tag{5.2.18}$$

where the coefficients $d_{1(*)}^{(j)}$'s and $d_{0(*)}^{(j)}$ are obtained by minimizing

$$\underset{j=1}{\overset{m}{\Sigma}} E(\theta_j - \underset{l=1}{\overset{m}{\Sigma}} d_l^{(j)} \underset{\sim}{\bar{x}}_l - d_0^{(j)})^2 \tag{5.2.19}$$

with respect to $d_l^{(j)}$'s and $d_0^{(j)}$. Now,

$$E(\bar{X}_1) = \nu, \tag{5.2.20}$$

$$E(\bar{X}_1^2) = n_1^{-1}\tau^2 + \sigma^2 + \theta^2 + \nu^2, \tag{5.2.21}$$

$$E(\bar{X}_1\bar{X}_{i'}) = \theta^2 + \nu^2 \quad (i \neq 1'), \tag{5.2.22}$$

$$E(\theta_j\bar{X}_j) = \sigma^2 + \theta^2 + \nu^2, \tag{5.2.23}$$

$$E(\theta_j\bar{X}_1) = \theta^2 + \nu^2 \quad (j \neq 1). \tag{5.2.24}$$

Using Theorem 2.1 of Goldstein (1975a) once again, it follows that for $1 \leq j \leq m$,

$$E(\theta_j|\underset{\sim}{\bar{X}}) = (1 \ \underset{\sim}{\bar{X}}')\underset{\sim}{D}^{-1}\ \underset{\sim}{b}, \tag{5.2.25}$$

where

$$\underset{\sim}{D} = \begin{bmatrix} 1 & \nu\underset{\sim}{1}_m' \\ \nu\underset{\sim}{1}_m & \underset{\sim}{C}+\theta^2\underset{\sim}{1}_m\underset{\sim}{1}_m' \end{bmatrix}$$

$$\underset{\sim}{b} = \begin{bmatrix} \nu \\ \theta^2\underset{\sim}{1}_m+\sigma^2\underset{\sim}{e}_m^{(j)} \end{bmatrix} \tag{5.2.26}$$

In the above

$$\underset{\sim}{C} = \mathrm{Diag}(n_1^{-1}\tau^2+\sigma^2,\ldots,n_m^{-1}\tau^2+\sigma^2)$$

$$= \sigma^2\mathrm{Diag}(n_1^{-1}R_1+1,\ldots,n_m^{-1}R_1+1), \ R_1 = \tau^2/\sigma^2.$$

$\underset{\sim}{e}_m^{(j)}$ is a m-component column vector with its jth component equal

[illegible] and [illegible] rest matrices. Write $B_j = R_1/(R_1+R_j)$. Then $\underset{\sim}{C} = \sigma^2 \mathrm{Diag}\{(1-B_j)^{-1},\ldots,(1-B_m)^{-1}\}$. Now, in order to invert $\underset{\sim}{D}$, we use the following matrix results given on page 33 of Rao (1973).

Result 5.2.1 When A, D are symmetric matrices such that the inverses which occur in the expressions exist, that

$$\begin{bmatrix} A & E \\ E' & D \end{bmatrix}^{-1} = \begin{bmatrix} A^{-1} + FE^{-1}F' & -FE^{-1} \\ -E^{-1}F' & E^{-1} \end{bmatrix}$$

where $E = D - E'A^{-1}E$, $F = A^{-1}E$.

Result 5.2.2. Let A be a nonsingular matrix, and U and V be two column vectors. Then

$$(A+UV')^{-1} = A^{-1} - \frac{(A^{-1}U)(V'A^{-1})}{1+V'A^{-1}U}$$

In our case $E = \underset{\sim}{C} + \sigma^2 \underset{\sim}{1}_m \underset{\sim}{1}_m'$ and $F = \sigma^2 \underset{\sim}{1}_m$.

Applying Result 5.2.2,

$$\underset{\sim}{E}^{-1} = \underset{\sim}{C}^{-1} - \frac{\sigma^2}{1+\sigma^2 \underset{\sim}{1}_m' \underset{\sim}{C}^{-1} \underset{\sim}{1}_m} \underset{\sim}{C}^{-1} \underset{\sim}{1}_m \underset{\sim}{1}_m' \underset{\sim}{C}^{-1}. \tag{5.2.27}$$

Now,

$$\underset{\sim}{1}_m' \underset{\sim}{C}^{-1} \underset{\sim}{1}_m = \sigma^{-2} \sum_{j=1}^{m} (1-B_j) \tag{5.2.28}$$

and

$$\underset{\sim}{C}^{-1} \underset{\sim}{1}_m \underset{\sim}{1}_m' \underset{\sim}{C}^{-1} = \sigma^{-4} \{((1-B_j)(1-B_{j'}))\}. \tag{5.2.29}$$

where $R_2 = \sigma^2/\sigma^2$, $\underset{\sim}{C}_1 = \mathrm{Diag}(1-B_1,\ldots,1-B_m)$ and

$\underset{\sim}{G}_2 = [(1-E_j)(1-E_{j'})]$. the matrix with its (j,j') element equal to $(1-E_j)(1-E_{j'})$. Now, it follows from (5.2.27)-(5.2.29)

$$E^{-1} = \sigma^{-2}\left(\underset{\sim}{G}_1 - \frac{R_2}{1+R_2\sum_{j=1}^{m}(1-E_j)}\,\underset{\sim}{G}_2\right). \tag{5.2.30}$$

Also,

$$\begin{aligned} &FE^{-1} \\ &= v\underset{\sim}{1}'\underset{\sim}{C}^{-1} - \frac{v\xi^2}{1+\xi^2\underset{\sim}{1}_m'\underset{\sim}{C}^{-1}\underset{\sim}{1}_m}\,\underset{\sim}{1}_m'\underset{\sim}{C}^{-1}\underset{\sim}{1}_m\underset{\sim}{1}_m'\underset{\sim}{C}^{-1} \\ &= v\sigma^{-2}(\underset{\sim}{1}_m - \underset{\sim}{E})' - \frac{v\xi^2}{1+R_2\sum_{j=1}^{m}(1-E_j)}\,\sigma^{-4}\sum_{j=1}^{m}(1-E_j)(\underset{\sim}{1}_m - \underset{\sim}{E})' \\ &= v\sigma^{-2}\left(1+R_2\sum_{j=1}^{m}(1-E_j)\right)^{-1}(\underset{\sim}{1}_m - \underset{\sim}{E})' \end{aligned} \tag{5.2.31}$$

and

$$\begin{aligned} &\underset{\sim}{E}E^{-1}\underset{\sim}{E} \\ &= v^2\underset{\sim}{1}_m'\underset{\sim}{C}^{-1}\underset{\sim}{1}_m - \frac{v^2\xi^2}{1+\xi^2\underset{\sim}{1}_m'\underset{\sim}{C}^{-1}\underset{\sim}{1}_m}\,\underset{\sim}{1}_m'\underset{\sim}{C}^{-1}\underset{\sim}{1}_m\underset{\sim}{1}_m'\underset{\sim}{C}^{-1}\underset{\sim}{1}_m \\ &= v^2\sigma^{-2}\sum_{j=1}^{m}(1-E_j) - \frac{v^2\xi^2}{1+\xi^2\sigma^{-2}\sum_{j=1}^{m}(1-E_j)}\,\sigma^{-4}\left[\sum_{j=1}^{m}(1-E_j)\right]^2 \\ &= v^2\sigma^{-2}\sum_{j=1}^{m}(1-E_j)\left(1+R_2\sum_{j=1}^{m}(1-E_j)\right)^{-1}. \end{aligned} \tag{5.3.32}$$

Using Result 5.2.1, it follows from (5.2.30)-(5.2.32),

$$\underset{\sim}{D}^{-1} = \begin{bmatrix} 1+v\sum_{j=1}^{m}(1-E_j)\left(1+R_2\sum_{j=1}^{m}(1-E_j)\right)^{-1} & -v\sigma^{-2}\left(1+R_2\sum_{j=1}^{m}(1-E_j)\right)^{-1}(\underset{\sim}{1}_m - \underset{\sim}{E})' \\ -v\sigma^{-2}\left(1+R_2\sum_{j=1}^{m}(1-E_j)\right)^{-1}(\underset{\sim}{1}_m - \underset{\sim}{E}) & \sigma^{-2}\left(\underset{\sim}{G}_1 - \frac{R_2}{1+R_2\sum_{j=1}^{m}(1-E_j)}\,\underset{\sim}{G}_2\right) \end{bmatrix} \tag{5.2.33}$$

where $v = v^2/s^2$. Using some heavy algebra once again, one gets

$$\underset{\sim}{D}^{-1}\underset{\sim}{b} = \begin{bmatrix} \nu B_j\left(1+R_2\Sigma_{j=1}^{m}(1-B_j)\right)^{-1} \\ R_2 B_j\left(1+R_2\Sigma_{j=1}^{m}(1-B_j)\right)^{-1}(\underset{\sim}{1}_m-\underset{\sim}{B})+(1-B_j)\underset{\sim}{e}_m^{(j)} \end{bmatrix} \tag{5.2.34}$$

Hence, for $1<j<m$,

$$\begin{aligned} E(\theta_j|\underset{\sim}{X}) &= \nu B_j\left(1+R_2\Sigma_{j=1}^{m}(1-B_j)\right)^{-1} \\ &\quad +R_2 B_j\left(1+R_2\Sigma_{j=1}^{m}(1-B_j)\right)^{-1}\Sigma_{j=1}^{m}(1-B_j)\bar{X}_j+(1-B_j)\bar{X}_j \\ &= B_j\left[\left(1+R_2\Sigma_{j=1}^{m}(1-B_j)\right)^{-1}\left(\nu+R_2\Sigma_{j=1}^{m}(1-B_j)\bar{X}_j\right)\right]+(1-B_j)\bar{X}_j \\ &= \bar{X}_j-B_j\left[\bar{X}_j-\left(1+R_2\Sigma_{j=1}^{m}(1-B_j)\right)^{-1}\left(\nu+R_2\Sigma_{j=1}^{m}(1-B_j)\bar{X}_j\right)\right] \\ &= \bar{X}_j - B_j(\bar{X}_j-u) \qquad \text{(say)}, \end{aligned} \tag{5.2.35}$$

where $u = \left(1+R_2\Sigma_{j=1}^{m}(1-B_j)\right)^{-1}\left(\nu+R_2\Sigma_{j=1}^{m}(1-B_j)\bar{X}_j\right)$.

For $m+1<j<M$,

$$E(\theta_j|\underset{\sim}{X}) = (1\ \underset{\sim}{X}')\underset{\sim}{D}^{-1}\begin{bmatrix} \nu \\ \underset{\sim}{0}_m \end{bmatrix} \tag{5.2.36}$$

Hence, writing $N = \Sigma_{j=1}^{M}N_j$ and $n = \Sigma_{j=1}^{m}n_j$, the Bayes estimator of $\gamma = N^{-1}\Sigma_{j=1}^{M}\Sigma_{i=1}^{N_j}X_{ji}$ is given by

$$e_B$$

$$= N^{-1}\{\Sigma_{j=1}^{m}\Sigma_{i=1}^{n_j}X_{ji}+\Sigma_{j=1}^{m}\Sigma_{i=n_j+1}^{N_j}E(X_{ji}\,|\,\underline{\chi})+\Sigma_{j=m+1}^{M}\Sigma_{i=1}^{N_j}E(X_{ji}\,|\,\underline{\chi})\}$$

$$= N^{-1}\{\Sigma_{j=1}^{m}n_j\bar{X}_j+\Sigma_{j=1}^{m}\Sigma_{i=n_j+1}^{N_j}E(\theta_j\,|\,\underline{\chi})+\Sigma_{j=m+1}^{M}\Sigma_{i=1}^{N_j}E(\theta_j\,|\,\underline{\chi})\}$$

$$= N^{-1}\{\Sigma_{j=1}^{m}n_j\bar{X}_j+\Sigma_{j=1}^{m}(N_j-n_j)(B_j u+(1-B_j)\bar{X}_j)+(N-\Sigma_{j=1}^{m}N_j)u\}. \qquad (5.2.37)$$

Now, writing $f_j = (N_j-n_j)/N_j$, $r_j = N_j/N$, one gets from (5.2.37),

$$e_B$$

$$= N^{-1}\{\Sigma_{j=1}^{m}(1-f_j)N_j\bar{X}_j+\Sigma_{j=1}^{m}N_j f_j(B_j u+(1-B_j)\bar{X}_j)+N(1-\Sigma_{j=1}^{m}r_j)u\}$$

$$= \Sigma_{j=1}^{m}(1-f_j)r_j\bar{X}_j+\Sigma_{j=1}^{m}r_j f_j(B_j u+(1-B_j)\bar{X}_j)+(1-\Sigma_{j=1}^{m}r_j)u$$

$$= \Sigma_{j=1}^{m}r_j(1-f_j B_j)\bar{X}_j+(1-\Sigma_{j=1}^{m}r_j(1-f_j B_j))u$$

$$= u+\Sigma_{j=1}^{m}r_j(1-f_j B_j)(\bar{X}_j-u). \qquad (5.2.38)$$

Back to the simultaneous estimation of strata means, the Bayes estimator of $\underline{\gamma} = (\gamma_1,\ldots,\gamma_m)'$ with $\gamma_k = N_k^{-1}\Sigma_{j=1}^{M_k}\Sigma_{i=1}^{N_{kj}}X_{kji}$ $(N_k = \Sigma_{j=1}^{M_k}N_{kj})$ is given by $e_{\underline{B}} = (e_B^{(1)},\ldots,e_B^{(k)})'$ with

$$e_B^{(k)} = u_k+\Sigma_{j=1}^{m_k}r_{kj}(1-f_{kj}B_{kj})(\bar{X}_{kj}-u_k), \qquad (5.2.39)$$

where $B_{kj} = B_1/(B_1+n_{kj})$ (recall that $B_1 = \tau^2/\sigma^2$),

$\bar{X}_{kj} = n_{kj}^{-1}\Sigma_{i=1}^{n_{kj}}X_{kji}$, $u_k = \left(1+B_2\Sigma_{j=1}^{m_k}(1-B_{kj})\right)^{-1}(u+B_2\Sigma_{j=1}^{m_k}(1-B_{kj})\bar{X}_{kj})$,

$f_{kj} = (N_{kj}-n_{kj})/N_{kj}$, and $r_{kj} = N_{kj}/N_k$.

Remark 5.2.1. There is an interesting interpretation of w_k. Since $R_2 = \delta^2/\sigma^2$, one can write

$$w_k = \{\delta^{-2}+\Sigma_{j=1}^{m_k}\sigma^{-2}(1-R_{kj})\}^{-1}\{\delta^{-2}v+\Sigma_{j=1}^{m_k}\sigma^{-2}(1-R_{kj})\bar{X}_{kj}\}. \qquad (5.2.40)$$

We may note that $v = E(\mu_k)$ and $\delta^2 = V(\mu_k)$. Also,

$$E[V(\bar{X}_{kj}|\mu_k)] = n_{kj}^{-1}\tau^2+\delta^2 = \sigma^2\{n_{kj}^{-1}n_{kj}+1\} = \sigma^2(1-R_{kj})^{-1} .$$

Hence from (5.2.40), w_k is a weighted average of the prior mean v and the sample means $\bar{X}_{k1},\ldots,\bar{X}_{km_k}$, where the weights are reciprocals of $V(\mu_k)$, $EV(\bar{X}_{k1}|\mu_k),\ldots,EV(\bar{X}_{km_k}|\mu_k)$. Also, it is clear from (5.2.39) that the Bayes estimator shrinks the sample means $\bar{X}_{kj}$'s to the pooled mean w_k.

Remark 5.2.2. For K=1, Ericson (1976) also obtains an estimator of the finite population mean without making any distributional assumption. But, his main emphasis is on Bayesian interpretation of some standard sampling theory results. Ericson's (1976) basic assumptions are different from (A)-(C), and so he ends up with a different estimator of the finite population mean.

Remark 5.2.3. Note that when $N_k = 1$ for all $k = 1,\ldots,K$, and $\delta^2 = 0$ so that $R_2 = \delta^2/\sigma^2 = 0$, and $w_k = v$, one gets

$$e_B^{(k)} = N_{k1}^{-1}[n_{k1}\bar{X}_{k1}+(N_{k1}-n_{k1})\{(1-R_{k1})\bar{X}_1+R_{k1}v\}] \qquad (5.2.41)$$

which is the same as the estimator considered in (2.2.1), where

the notation μ is used for ν. In fact, this estimator agrees with the Bayes estimator of the finite population mean as obtained earlier by Ericson (1969a, 1969b).

Remark 5.2.4. Scott and Smith (1969) considered two-stage estimation of the finite population mean in the case of a single stratum under the normality assumption of the prior. In order to make our results comparable to theirs, we consider the case when $\delta^2 = 0$. Then $u = v$, and the expression given in (5.2.38) simplifies to

$$e_B = v + \Sigma_{j=1}^{m} r_j (1-f_j B_j)(\bar{X}_j - v). \tag{5.2.42}$$

Using our notations, writing $\tilde{v} = \Sigma_{j=1}^{m}(1-B_j)\bar{X}_j / \Sigma_{j=1}^{m}(1-B_j)$, and correcting an obvious typographical error in equation (4) of Scott and Smith, it follows that with their $e_{ji} = N^{-1}$ (this is our N which is different from the N of Scott and Smith), the Bayes estimator of the finite population mean proposed by Scott and Smith is

$$\begin{aligned} & e_B^* \\ &= N^{-1}\left[\Sigma_{j=1}^{m}\Sigma_{i=1}^{n_j} X_{ji} + \Sigma_{j=1}^{m}(N_j - n_j)(1-B_j)\bar{X}_j + \tilde{v}\{\Sigma_{j=1}^{m}(N_j - n_j)B_j + \Sigma_{j=m+1}^{M} N_j\}\right] \\ &= \Sigma_{j=1}^{m} r_j (1-f_j)\bar{X}_j + \Sigma_{j=1}^{m} r_j f_j (1-B_j)\bar{X}_j + \tilde{v}\{\Sigma_{j=1}^{m} r_j (1-f_j)B_j + 1 - \Sigma_{j=1}^{m} r_j\} \\ &= \tilde{v} + \Sigma_{j=1}^{m} r_j (1-f_j B_j)(\bar{X}_j - \tilde{v}) \end{aligned} \tag{5.2.43}$$

One can see the closeness of the two expressions given in (5.2.43) and (5.2.43), the only difference being that in (5.2.43), v is replaced by $\tilde{v}$. We shall see in the next section that e_{ij}^{*} can indeed be motivated from an empirical Bayes point of view.

5.3 Empirical Bayes Estimation

First assume that all the variance components τ^2, α^2, and δ^2 are known, but the parameter v is unknown, and needs to be estimated from the data. Writing $\bar{X}_k' = (\bar{X}_{k1},\ldots,\bar{X}_{km_k})$ and

$$\begin{aligned} C_k &= \mathrm{Diag}(n_{k1}^{-1}\tau^2+\alpha^2,\ldots,n_{km_k}^{-1}\tau^2+\alpha^2) \\ &= \alpha^2\mathrm{Diag}((1-B_{k1})^{-1},\ldots,(1-B_{km_k})^{-1}). \end{aligned}$$

it follows that $E(\bar{X}_k) = v1_{m_k}$ and $V(\bar{X}_k) = C_k+\delta^2 1_{m_k}1_{m_k}'$. In the absence of any distributional assumption on the prior, we find the BLUE of v which is obtained by minimizing

$$Q = \sum_{k=1}^{K}(\bar{X}_k-v1_{m_k})'(C_k+\delta^2 1_{m_k}1_{m_k}')^{-1}(\bar{X}_k-v1_{m_k}). \qquad (5.3.1)$$

As in (5.2.33),

$$\begin{aligned} &(C_k+\delta^2 1_{m_k}1_{m_k}')^{-1} \\ &= \alpha^{-2}[\mathrm{Diag}(1-B_{k1},\ldots,1-B_{km_k}) \\ &\qquad -B_2(1+B_2\textstyle\sum_{j=1}^{m}(1-B_{kj}))^{-1}((1-B_{kj})(1-B_{kj'}))]. \end{aligned}$$

[illegible]

$$e^2Q = \Sigma_{k=1}^{K}\Sigma_{j=1}^{m_k}(\bar{X}_{kj}-v)^2(1-B_{kj}) - \Sigma_{k=1}^{K}\{\Sigma_{j=1}^{m_k}(\bar{X}_{kj}-v)(1-B_{kj})\}^2 d_k R_2, \tag{5.3.2}$$

where $d_k = \{1+R_2\Sigma_{j=1}^{m_k}(1-B_{kj})\}^{-1}$. Now, differentiating e^2Q with respect to v, and equating the derivative to zero, one gets

$$0 = \frac{\partial}{\partial v}(e^2Q) =$$

$$-2\Sigma_{k=1}^{K}\Sigma_{j=1}^{m_k}(\bar{X}_{kj}-v)(1-B_{kj})+2\Sigma_{k=1}^{K}d_k\{\Sigma_{j=1}^{m_k}(\bar{X}_{kj}-v)(1-B_{kj})\}\Sigma_{j=1}^{m_k}(1-B_{kj})R_2. \tag{5.3.3}$$

Noting that $R_2 d_k \Sigma_{j=1}^{m_k}(1-B_{kj}) - 1 = -d_k$, it follows from (5.3.3) that the BLUE of v is given by

$$\hat{\hat{v}} = \{\Sigma_{k=1}^{K}[\Sigma_{j=1}^{m_k}(1-B_{kj})]d_k\}^{-1}\{\Sigma_{k=1}^{K}[\Sigma_{j=1}^{m_k}(1-B_{kj})\bar{X}_{kj}]d_k\}$$
$$- \{\Sigma_{k=1}^{K}d_k\Sigma_{j=1}^{m_k}(1-B_{kj})\}^{-1}\{\Sigma_{k=1}^{K}d_k\Sigma_{j=1}^{m_k}(1-B_{kj})\bar{X}_{kj}\}. \tag{5.3.4}$$

In addition, if one assumes normality, $\hat{\hat{v}}$ is also the MLE of v. The empirical Bayes estimator of $\underset{\sim}{\chi}$ is then

$$\underset{\sim}{e}_{EB}^{*} = (e_{EB}^{1(*)},\ldots,e_{EB}^{K(*)})', \text{ where}$$

$$e_{EB}^{k(*)} = \hat{u}_k + \Sigma_{j=1}^{m_k}[r_{kj}(1-f_{kj}B_{kj})(\bar{X}_{kj}-\hat{u}_k), \tag{5.3.5}$$

$\hat{u}_k$ being obtained replacing v by $\hat{\hat{v}}$ in u_k.

Remark 5.3.1 The estimator $\hat{\theta}$ has an interesting interpretation. Writing $\bar{X}_{wk} = \Sigma_{j=1}^{m_k}(1-g_{kj})\bar{X}_{kj}/\Sigma_{j=1}^{m_k}(1-g_{kj})$ and $c_k = d_k\Sigma_{j=1}^{m_k}(1-g_{kj})$, it follows from (5.3.4) that $\hat{\theta}$ can be alternatively expressed as

$$\hat{\theta} = \Sigma_{k=1}^{K} c_k \bar{X}_{wk} / \Sigma_{k=1}^{K} c_k \tag{5.3.6}$$

Note that $\bar{X}_{wk}$ is itself a weighted average of the sample means $\bar{X}_{k1},\ldots,\bar{X}_{km_k}$ of the first stage units. This is because as derived in Remark 5.2.1, $E[V(\bar{X}_{kj}|\mu_k)] = \sigma^2(1-g_{kj})^{-1}$. Thus, $\bar{X}_{wk}$ is a weighted average of the $\bar{X}_{kj}$'s with weights inversely proportional to $E[V(\bar{X}_{kj}|\mu_k)]$. Moreover,

$$\begin{aligned} V(\bar{X}_{wk}) &= [\Sigma_{j=1}^{m_k}(1-g_{kj})]^{-2}[\Sigma_{j=1}^{m_k}(1-g_{kj})^2 V(\bar{X}_{kj}) \\ &\qquad + \sum_{1\le j\ne j'\le m_k}(1-g_{kj})(1-g_{kj'})\mathrm{Cov}(\bar{X}_{kj},\bar{X}_{kj'})\} \\ &= [\Sigma_{j=1}^{m_k}(1-g_{kj})]^{-2}[\Sigma_{j=1}^{m_k}(1-g_{kj})^2\{\sigma^2(1-g_{kj})^{-1}+\delta^2\} \\ &\qquad + \sum_{1\le j\ne j'\le m_k}(1-g_{kj})(1-g_{kj'})\delta^2] \\ &= \sigma^2[\Sigma_{j=1}^{m_k}(1-g_{kj})]^{-1} + \delta^2 \\ &= \sigma^2[\Sigma_{j=1}^{m_k}(1-g_{kj})]^{-1}\{1+R_2\Sigma_{j=1}^{m_k}(1-g_{kj})\} \\ &= \sigma^2[d_k\Sigma_{j=1}^{m_k}(1-g_{kj})]^{-1} = \sigma^2 c_k^{-1}. \end{aligned} \tag{5.3.7}$$

Thus, it follows from (5.3.6) and (5.3.7) that $\hat{\theta}$ is a weighted average of $\bar{X}_{wk}$'s with weights inversely proportional to $V(\bar{X}_{wk})$.

Remark 5.3.2. Note that for one stratum case, since $s^2 = 0$ we have $R_2 = 0$, $\hat{\theta} = \tilde{v}$. Then, it agrees with the estimator e_B^* of Scott and Smith (1969). It is interesting to note that Scott and Smith had an alternative derivation of $\tilde{v}$. They assumed that conditional on the θ_j's, X_{j1}'s were iid $N(\theta_j, \tau^2)$, while conditional on v, θ_j's are iid $N(v, \sigma^2)$, and v is uniform $(-\infty, \infty)$. This is yet another example where an estimator obtained by hierarchial Bayesian analysis or Bayes empirical Bayes analysis also agrees with an estimator obtained by classical empirical Bayes analysis. This is not surprising though since the estimator $\tilde{v}$ is indeed a generalized version of the finite population analogue of the classic estimator of Lindley and Smith (1972) which is well-known to be both classical empirical Bayes and Bayes empirical Bayes.

So far the variance components τ^2, σ^2, and s^2 and accordingly $R_1 = \tau^2/\sigma^2$, and $R_2 = s^2/\sigma^2$ are assumed to be known. For the most practical situations, however, these parameters are unknown, and need to be estimated from the data. To estimate these parameters, we define

$$SS_1 = \sum_{k=1}^{K} \sum_{j=1}^{m_k} \sum_{i=1}^{n_{kj}} (X_{kji} - \bar{X}_{kj})^2, \tag{5.3.8}$$

$$SS_2 = \sum_{k=1}^{K} \sum_{j=1}^{m_k} n_{kj} (\bar{X}_{kj} - \bar{X}_k)^2, \tag{5.3.9}$$

$$SS_3 = \sum_{k=1}^{K} n_k (\bar{X}_k - \bar{\bar{X}})^2, \tag{5.3.10}$$

where $\bar{X}_{kj} = n_{kj}^{-1} \sum_{i=1}^{n_{kj}} X_{kji}$, $n_k = \sum_{j=1}^{m_k} n_{kj}$ = total number of secondary

unit in the kth stratum, $\bar{X}_k = n_k^{-1}\sum_{j=1}^{m_k} n_{kj}\bar{X}_{kj}$, and $\bar{\bar{X}} = n^{-1}\sum_{k=1}^{K} n_k\bar{X}_k$, where $n = \sum_{k=1}^{K} n_k$. Write $X_{kji} = Z_{kji} + \theta_{kj}$ and define correspondingly $\tilde{Z}_{kj}$, $\bar{Z}_k$, $\bar{\bar{Z}}$ so that

$$SS_1 = \sum_{k=1}^{K}\sum_{j=1}^{m_k}\sum_{i=1}^{n_{kj}} (Z_{kji}-\bar{Z}_{kj})^2, \tag{5.3.11}$$

$$SS_2 = \sum_{k=1}^{K}\sum_{j=1}^{m_k} n_{kj}[(\bar{Z}_{kj}-\bar{Z}_k) + (\theta_{kj}-\bar{\theta}_k)]^2, \tag{5.3.12}$$

$$SS_3 = \sum_{k=1}^{K} n_k[(\bar{Z}_k-\bar{\bar{Z}}) + (\bar{\theta}_k-\bar{\bar{\theta}})]^2. \tag{5.3.13}$$

We prove the following lemma which motivates the estimation of the variance components.

Lemma 5.3.1 Under the model

$$E(SS_1) = (n-m)\tau^2, \text{ where } m = \sum_{k=1}^{K} m_k. \tag{5.3.14}$$

$$E(SS_2) = (n - \sum_{k=1}^{K} n_k^{-1}\sum_{j=1}^{m_k} n_{kj}^2)\sigma^2 + (m-K)\tau^2, \tag{5.3.15}$$

$$E(SS_3) = (n - n^{-1}\sum_{k=1}^{K} n_k^2)\sigma^2 + [\sum_{k=1}^{K}(n_k^{-1}-n^{-1})\sum_{j=1}^{m_k} n_{kj}^2]\sigma^2 + (K-1)\tau^2. \tag{5.3.16}$$

Proof of the Lemma. Note that conditional on θ_{kj}'s, Z_{kji}'s are independent with $E(Z_{kji}|\theta_{kj}) = 0$ and $E(Z_{kji}^2|\theta_{kj}) = \nu_2(\theta_{kj})$ for all $i=1,\ldots,n_{kj}$, $j=1,\ldots,m_k$, and $k=1,\ldots,K$. Also, marginally θ_{kj}'s are identically distributed so that

$$E\nu_2(\theta_{kj}) = E\nu_2(\theta_{11}) = \tau^2$$

for all $j=1,\ldots,m_k$ and $k=1,\ldots,K$.

Now, using Lemma 2.3.1, one gets from (5.3.11),

$$\begin{aligned} E(SS_1) &= \sum_{k=1}^{K}\sum_{j=1}^{m_k} E\{E[\sum_{i=1}^{n_{kj}}(Z_{kji}-\bar{Z}_{kj})^2|\theta_{kj}'s]\} \\ &= \sum_{k=1}^{K}\sum_{j=1}^{m_k} E\{\sum_{i=1}^{n_{kj}}(1-a_{kj}^{-1})u_2(\theta_{kj})\} \\ &= \sum_{k=1}^{K}\sum_{j=1}^{m_k}(n_{kj}-1)\tau^2 \\ &= (n-m)\tau^2. \end{aligned}$$

This proves (5.3.14)

From (5.3.12), using $E(\bar{Z}_{kj}-\bar{\bar{Z}}_k|\theta_{kj}) = 0$ for all $j=1,\ldots,m_k$ and $k=1,\ldots,K$, one gets

$$\begin{aligned} E(SS_2) = \sum_{k=1}^{K} E\{&E[\sum_{j=1}^{m_k} n_{kj}(\bar{Z}_{kj}-\bar{\bar{Z}}_k)^2|\theta_{kj}'s] \\ &+\sum_{j=1}^{m_k} n_{kj}(\theta_{kj}-\bar{\bar{\theta}}_k)^2\}. \end{aligned} \tag{5.3.17}$$

Now,

$$\sum_{j=1}^{m_k} n_{kj}(\bar{Z}_{kj}-\bar{\bar{Z}}_k)^2 = \bar{\underset{\sim}{Z}}_k'(\underset{\sim}{N}_k - a_k^{-1}\underset{\sim}{n}_{(k)}\underset{\sim}{n}_{(k)}')\bar{\underset{\sim}{Z}}_k, \tag{5.3.18}$$

where

$$\bar{\underset{\sim}{Z}}_k = (\bar{Z}_{k1},\ldots,\bar{Z}_{km_k})',$$

$$\underset{\sim}{N}_k = \mathrm{Diag}(n_{k1},\ldots,n_{km_k}),$$

$$\underset{\sim}{n}_{(k)} = (n_{k1},\ldots,n_{km_k})'.$$

Note that conditional on θ_{kj}'s, $\bar{Z}_{kj}$'s are independent with

$$E(\bar{Z}_{kj}|\theta_{kj}) = 0 \text{ and } V(\bar{Z}_{kj}|\theta_{kj}) = n_{kj}^{-1}\mu_2(\theta_{kj}) \qquad (5.3.19)$$

for all $j=1,\ldots,m_k$ and $k=1,\ldots,K$.

Now, using (5.3.18), (5.3.19) and Lemma 2.3.1, one gets

$$E[\Sigma_{j=1}^{m_k} n_{kj}(\bar{Z}_{kj}-\bar{\bar{Z}}_k)^2 | \theta_{kj}'s]$$
$$= \Sigma_{j=1}^{m_k}(n_{kj} - n_k^{-1}n_{kj}^2)n_{kj}^{-1}\mu_2(\theta_{kj}). \qquad (5.3.20)$$

Writing $\bar{Z}_{kj} = Y_{kj} + \nu$, $\bar{Y}_k = n_k^{-1}\Sigma_{j=1}^{m_k} n_{kj}Y_{kj}$, one gets

$$\Sigma_{j=1}^{m_k} n_{kj}(\bar{Z}_{kj}-\bar{\bar{Z}}_k)^2 = \Sigma_{j=1}^{m_k} n_{kj}(Y_{kj}-\bar{Y}_k)^2, \qquad (5.3.21)$$

Note that conditional on μ_k's, Y_{kj}'s are independent with

$$E(Y_{kj}|\mu_k) = 0 \text{ and } V(Y_{kj}|\mu_k) = m_2(\mu_k). \qquad (5.3.22)$$

Applying Lemma 2.3.1 once again, one gets

$$E[\Sigma_{j=1}^{m_k} n_{kj}(Y_{kj}-\bar{Y}_k)^2 | \mu_k]$$
$$= (n_k - n_k^{-1}\Sigma_{j=1}^{m_k} n_{kj}^2)m_2(\mu_k). \qquad (5.3.23)$$

From (5.3.17), (5.3.20), and (5.3.23), (5.3.15) follows.

Using the fact that $E(\bar{Z}_k - \bar{\bar{Z}}|\theta_{kj}'s) = 0$, it follows from (5.3.13) that

$$E(SS_3) = E[\Sigma_{k=1}^{K} n_k(\bar{Z}_k - \bar{\bar{Z}})^2] + E[\Sigma_{k=1}^{K} n_k(\bar{\theta}_k - \bar{\theta})^2]. \qquad (5.3.24)$$

Noting that $EV(\bar{z}_k|\theta_{kj}\text{'s}) = n_k^{-1}\tau^2$, applying Lemma 2.3.1, one gets

$$E\{\Sigma_{k=1}^{K} n_k(\bar{z}_k-\bar{\bar{z}})^2\} = \Sigma_{k=1}^{K}(n_k-n^{-1}n_k^2)n_k^{-1}\tau^2$$

$$= (K-1)\tau^2. \qquad (5.3.25)$$

Since $\bar{Y}_k$'s are independent with $E(\bar{Y}_k) = 0$ and

$$V(\bar{Y}_k) = EV(\bar{Y}_k) + VE(\bar{Y}_k|\mu_k) = \{n_k^{-2}\Sigma_{j=1}^{m_k} n_{kj}^2\}\sigma^2 + d^2, \qquad (5.2.26)$$

applying Lemma 2.3.1 once again, one gets

$$E[\Sigma_{k=1}^{K} n_k(\bar{\theta}_k-\bar{\bar{\theta}})^2] = E[\Sigma_{k=1}^{K} n_k(\bar{Y}_k-\bar{Y})^2]$$

$$= \Sigma_{k=1}^{K}\{n_k-n^{-1}n_k^2\}[\{n_k^{-2}\Sigma_{j=1}^{m_k} n_{kj}^2\}\sigma^2+\delta^2]. \qquad (5.3.27)$$

Equation (5.3.16) now follows from (5.3.25) and (5.3.27).

Define now

$$MS_1 = SS_1/(n-m),\ MS_2 = SS_2/(m-K) \text{ and } MS_3 = SS_3/(K-1). \qquad (5.3.28)$$

It follows then from (5.3.14)-(5.3.16) that

$$E(MS_1) = \tau^2, \qquad (5.3.29)$$

$$E(MS_2) = \tau^2 + g_1\sigma^2, \qquad (5.3.30)$$

$$E(MS_3) = \tau^2 + g_2\sigma^2 + g_3\delta^2, \qquad (5.3.31)$$

where

$$g_1 = (m-K)^{-1}(n-\Sigma_{k=1}^{K} n_k^{-1}\Sigma_{j=1}^{m_k} n_{kj}^2),$$

$$g_2 = (K-1)^{-1} \sum_{k=1}^{K} (n_k^{-1} - n^{-1}) \sum_{j=1}^{m_k} n_{kj}^2$$

and

$$g_3 = (K-1)^{-1}(n-n^{-1} \sum_{k=1}^{K} n_k^2).$$

Now, observing that

$$E[(MS_2 - MS_1)g_1^{-1}] = \sigma^2$$

and

$$E[\{MS_3 - MS_1 - g_2 g_1^{-1}(MS_2 - MS_1)\}g_3^{-1}] = \theta^2,$$

one estimates R_1 and R_2 by

$$\hat{R}_1 = \max\{0, (\frac{MS_2}{MS_1} - 1)^{-1} g_1\}, \tag{5.3.32}$$

$$\hat{R}_2 = \max[0, \{g_1(MS_3 - MS_1)(MS_2 - MS_1)^{-1} - g_2\}g_3^{-1}\}. \tag{5.3.33}$$

Now, let $\hat{B}_{kj} = \hat{R}_1(\hat{R}_1 + n_{kj})^{-1}$, $\hat{d}_k = (1 + \hat{R}_2 \sum_{j=1}^{m_k}(1-\hat{B}_{kj}))^{-1}$,

$\hat{v} = \{\sum_{k=1}^{K} \sum_{j=1}^{m_k} \hat{B}_{kj}(1-\hat{B}_{kj})\hat{d}_k\}^{-1}\{\sum_{k=1}^{K} \hat{d}_k r_{j=1}^{m_k}(1-\hat{B}_{kj})\bar{X}_{kj}\}$ and

$$\hat{u}_k = \hat{d}_k\{\hat{v} + \hat{R}_2 \sum_{j=1}^{m_k}(1-\hat{B}_{kj})\bar{X}_{kj}\}. \tag{5.3.34}$$

The empirical Bayes estimator of χ is now given by

$\hat{\chi}_{EB} = (e_{EB}^{(1)}, \ldots, e_{EB}^{(k)})$, where

$$e_{EB}^{(k)} = \hat{u}_k + \sum_{j=1}^{m_k} r_{kj}(1-f_{kj}\hat{B}_{kj})(\bar{X}_{kj} - \hat{u}_k). \tag{5.3.35}$$

5.4 RSL Behavior

In this section, we compare the performance of e^{*}_{EB} and e_{EB} with another estimator e_0 of $\bar{Y}$, where $e_0 = (e_0^{(1)}, \ldots, e_0^{(K)})$ with

$$e_0^{(k)} = N_k^{-1}\Big[\sum_{j=1}^{m_k}\sum_{i=1}^{n_{kj}} y_{kji} + \sum_{j=1}^{m_k}\sum_{i=n_{kj}+1}^{N_{kj}} \bar{y}_{kj} + \sum_{j=m_k+1}^{M_k}\sum_{i=1}^{N_{kj}} \{n_k^{-1}\sum_{j=1}^{m_k} e_{kj}\bar{x}_{kj}\}\Big]. \tag{5.4.1}$$

The above estimator was also considered in Royall (1976). The motivation of the estimator $e_0^{(k)}$ is that any unobserved secondary unit within a chosen primary unit is estimated by the sample mean of the corresponding primary unit, whereas any unobserved secondary unit within a primary unit which has not been selected is estimated by the pooled mean over all the primary units. After simplification, one gets from (5.4.1)

$$\begin{aligned} e_0^{(k)} &= N_k^{-1}\Big[\sum_{j=1}^{m_k}\{N_{kj}+n_k^{-1}e_{kj}(N_k-\sum_{j=1}^{m_k}N_{kj})\}\bar{x}_{kj} \\ &= \sum_{j=1}^{m_k}\Big[\{r_{kj}+w_{kj}(1-\sum_{j=1}^{m_k}r_{kj})\}\bar{x}_{kj}\Big], \end{aligned} \tag{5.4.2}$$

where $w_{kj} = n_{kj}/n_k$. The estimator $e_0^{(k)}$ is not the same as the usual design unbiased estimator of $\bar{Y}$, as given, for example, in Cochran (1977) or Raj (1968) except in the balanced case when all the primary units are of same size, and subsamples of same size are drawn from every selected primary unit. However, $e_0^{(k)}$ seems to be as sensible as some of the estimators proposed in the literature.

First we compare the Bayes risk performance of $\underset{\sim}{\delta}^*_{EB}$ and $\underset{\sim}{\delta}_{EB}$ with $\underset{\sim}{\delta}_0$. Denote by ξ the prior given in (A)-(C). We denote by $r(\xi,\underset{\sim}{\delta})$ the integrated Bayes risk of an estimator $\underset{\sim}{\delta}$ of $\underset{\sim}{\gamma}$ when the prior is ξ. From Lemma 2.4.1, it follows that

$$RSL(\xi;\underset{\sim}{\delta}^*_{EB},\underset{\sim}{\delta}_0) = E\|\underset{\sim}{\delta}^*_{EB}-\underset{\sim}{\delta}_B\|^2/E\|\underset{\sim}{\delta}_0-\underset{\sim}{\delta}_B\|^2, \tag{5.4.3}$$

where $\|.\|$ denotes the square of the Euclidean norm.

From (5.2.39) and (5.3.5), one gets

$$E\|\underset{\sim}{\delta}^*_{EB}-\underset{\sim}{\delta}_B\|^2 = E[\Sigma_{k=1}^K(1-\Sigma_{j=1}^{m_k}\tau_{kj}(1-f_{kj}B_{kj}))^2(\overset{*}{u}_k-u_k)^2]. \tag{5.4.4}$$

Note that by definition,

$$E(\overset{*}{u}_k-u_k)^2 = d_k^2E(\hat{\hat{v}}-v)^2 = d_k^2V(\hat{\hat{v}}). \tag{5.4.5}$$

Using the representation (5.3.6) for $\hat{\hat{v}}$, (5.3.7), and the independence of $\bar{X}_{ik}$'s, it follows from (5.4.5) that

$$V(\hat{\hat{v}}) = a^2(\Sigma_{k=1}^K c_k)^{-1} = a^2\{\Sigma_{k=1}^K d_k\Sigma_{j=1}^{m_k}(1-B_{kj})\}^{-1}. \tag{5.4.6}$$

Hence, from (5.4.4)-(5.4.6),

$$E\|\underset{\sim}{\delta}^*_{EB}-\underset{\sim}{\delta}_B\|^2 = a^2\{\Sigma_{k=1}^K d_k\Sigma_{j=1}^{m_k}(1-B_{kj})\}^{-1}\Sigma_{k=1}^K d_k^2(1-\Sigma_{j=1}^{m_k}\tau_{kj}(1-f_{kj}B_{kj}))^2. \tag{5.4.7}$$

Next to compute $E\|\underset{\sim}{\delta}_0-\underset{\sim}{\delta}_B\|^2$, first observe that after some simplifications, one gets

$$\delta_0^{(k)}-\delta_B^{(k)} = \Sigma_{j=1}^{m_k}v_{kj}(\bar{X}_{kj}-u_k), \tag{5.4.8}$$

where $v_{kj} = r_{kj}f_{kj}\bar{x}_{kj}+w_{kj}(1-\sum_{j=1}^{m_k} r_{kj})$. In deriving (5.4.8), we have used the fact that $\sum_{j=1}^{m_k} w_{kj} = 1$. Since $E(\bar{x}_{kj}) = E(u_k) = \nu$, it follows from (5.4.8) that

$$E(e_0^{(k)}-e_s^{(k)})^2$$

$$= \sum_{j=1}^{m_k} v_{kj}^2 V(\bar{x}_{kj}-u_k)+\sum_{1\le j\ne j'\le m_k} v_{kj}v_{kj'}Cov(\bar{x}_{kj}-u_k,\bar{x}_{kj'}-u_k). \qquad (5.4.9)$$

Next, using $V(\bar{x}_{kj}) = \sigma^2[(1-R_{kj})^{-1}+R_2]$, $V(u_k) = \sigma^2R_2(1-d_k)$, $Cov(\bar{x}_{kj},u_k) = \sigma^2R_2$ and $Cov(\bar{x}_{kj},\bar{x}_{kj'}) = \sigma^2R_2$ $(j\ne j')$, one gets from (5.4.9),

$$E(e_0^{(k)}-e_s^{(k)})^2$$

$$= \sigma^2[\sum_{j=1}^{m_k} v_{kj}^2[(1-R_{kj})^{-1}-R_2d_k]-R_2d_k\sum_{1\le j\ne j'\le m_k} v_{kj}v_{kj'}]$$

$$= \sigma^2[\sum_{j=1}^{m_k} v_{kj}^2(1-R_{kj})^{-1}-R_2d_k(\sum_{j=1}^{m_k} v_{kj})^2]. \qquad (5.4.10)$$

Hence, from (5.4.10),

$$E\|e_0-e_s\|^2 = \sigma^2[\sum_{k=1}^{K}[\sum_{j=1}^{m_k} v_{kj}^2(1-R_{kj})^{-1}-R_2d_k(\sum_{j=1}^{m_k} v_{kj})^2]]. \qquad (5.4.11)$$

Hence, from (5.4-3), (5.4.7) and (5.4.11), it follows that

$$RSL(\bar{y}_{EB}^{*},\bar{e}_0)$$

$$= [\{\sum_{k=1}^{K} d_k\sum_{j=1}^{m_k}(1-R_{kj})\}^{-1}\sum_{k=1}^{K} d_k^2(\sum_{j=1}^{m_k} v_{kj})^2]$$

$$+ [\sum_{k=1}^{K}\sum_{j=1}^{m_k} v_{kj}^2(1-R_{kj})^{-1}-R_2\sum_{k=1}^{K} d_k(\sum_{j=1}^{m_k} v_{kj})^2], \qquad (5.4.12)$$

where we have used the fact that $\sum_{j=1}^{m_k} v_{kj} = 1-\sum_{j=1}^{m_k} r_{kj}(1-f_{kj}R_{kj})$.

The first main theorem of this section gives the asymptotic convergence of $RSL(\xi;\underset{\sim}{e}^*_{EB},\underset{\sim}{e}_0)$.

Theorem 5.4.1 Assume that

(i) $\inf_{1\le k\le K} \inf_{1\le j\le m_k} n_{kj} > 1$,

(ii) $\sup_{1\le k\le K} n_k < C$,

and

(iii) $\inf_{1\le k\le K} M_k/N_k = p > 0$.

Then

$$RSL(\xi;\underset{\sim}{e}^*_{EB},\underset{\sim}{e}_0) = O(K^{-1}) \text{ as } K \to \infty. \tag{5.4.13}$$

Proof. Since $1-B_{kj} = a_{kj}/(R_1+n_{kj}) \uparrow$ in n_{kj}, using assumption (i), one gets $1-B_{kj} \ge (R_1+1)^{-1}$. Hence, from (5.4.7),

$$\begin{aligned} K^{-1}E\|\underset{\sim}{e}^*_{EB}-\underset{\sim}{e}_B\|^2 &\le K^{-1}a^2(R_1+1)(\Sigma_{k=1}^K n_k d_k)^{-1}(\Sigma_{k=1}^K d_k^2) \\ &\le K^{-1}a^2(R_1+1)(\Sigma_{k=1}^K \xi_k)^{-1}(\Sigma_{k=1}^K d_k) \\ &= K^{-1}a^2(R_1+1) \end{aligned} \tag{5.4.14}$$

Next, from (5.4.11), using $R_2 d_k \Sigma_{j=1}^{m_k}(1-B_{kj}) = 1-\delta_k$, it follows that

$$\begin{aligned} &K^{-1}E\|\underset{\sim}{e}_0-\underset{\sim}{e}_B\|^2 \\ &= K^{-1}a^2\Sigma_{k=1}^K[\Sigma_{j=1}^{m_k} v_{kj}^2(1-B_{kj})^{-1}-(1-d_k)(\Sigma_{j=1}^{m_k} v_{kj})^2\{\Sigma_{j=1}^{m_k}(1-B_{kj})\}^{-1}], \end{aligned} \tag{5.4.15}$$

Next, putting

$$v_{kj}(1-\varepsilon_{kj})^{-1/2} = y_{kj},\ (1-\varepsilon_{kj})^{1/2}\{\Sigma_{j=1}^{m_k}(1-\varepsilon_{kj})\}^{-1/2} = z_{kj},$$

and using the Schwarz inequality

$$(\Sigma_{j=1}^{m_k} y_{kj}z_{kj})^2 \le (\Sigma_{j=1}^{m_k} y_{kj}^2)(\Sigma_{j=1}^{m_k} z_{kj}^2),$$

one gets

$$\Sigma_{j=1}^{m_k} v_{kj}^2(1-\varepsilon_{kj})^{-1} \ge (\Sigma_{j=1}^{m_k} v_{kj})^2\{\Sigma_{j=1}^{m_k}(1-\varepsilon_{kj})\}^{-1}.$$

Hence, from (5.4.15), one gets

$$\begin{aligned} K^{-1}\mathrm{E}\|\xi_0-\xi_\beta\|^2 &\ge K^{-1}\alpha^2\Sigma_{k=1}^{K} d_k(\Sigma_{j=1}^{m_k} v_{kj})^2\{\Sigma_{j=1}^{m_k}(1-\varepsilon_{kj})\}^{-1} \\ &\ge K^{-1}\rho^2\Sigma_{k=1}^{K} d_k(\Sigma_{j=1}^{m_k} v_{kj})^2 m_k^{-1}. \end{aligned} \tag{5.4.16}$$

Note that $d_k = (1+R_2\Sigma_{j=1}^{m_k}(1-\varepsilon_{kj}))^{-1} \ge (1+R_2 m_k)^{-1}$, and using assumption (ii) $m_k \le C$. Moreover, using assumption (iii),

$$\begin{aligned} &\Sigma_{j=1}^{m_k} v_{kj} = 1-\Sigma_{j=1}^{m_k}\tau_{kj}(1-\varepsilon_{kj}B_{kj}) \ge 1-\Sigma_{j=1}^{m_k}\tau_{kj} \\ &= N_k^{-1}\Sigma_{j=m_k+1}^{M_k} N_{kj} \ge (M_k-m_k)N_k^{-1} \ge (M_k-m_k M_k^{-1})p \ge (1-m_k(m_k+1)^{-1})p \\ &= p(m_k+1)^{-1} \ge p(C+1)^{-1}. \end{aligned}$$

Now, from (5.4.16), we have

$$K^{-1}\mathrm{E}\|\xi_0-\xi_\beta\|^2 \ge K^{-1}\rho^2(1+R_2C)^{-1}p^2(C+1)^{-2}C^{-1}K$$

$$= \sigma^2\beta^2(C+1)^{-2}(C(1+R_2C))^{-1}. \qquad (5.4.17)$$

Combining (5.4.14) and (5.4.17), one gets (5.4.13).

Remark 5.4.1 It follows from (5.4.14) that the empirical Bayes estimator $\underset{\sim}{a}^*_{EB}$ is asymptotically optimal in the sense of Robbins (1955).

Next we consider the situation when the variance components are unknown, and the empirical Bayes estimator $\underset{\sim}{a}_{EB}$ is used. We first show that MS_1, MS_2 and MS_3 are consistent estimators of their expectations under certain conditions. Recall the definitions of Z_{kji}'s, $\bar{Z}_{kj}$'s, $\bar{Z}_k$'s and $\bar{Z}$ as given earlier (after (5.3.10)). Let

$$\mu_4(\theta_{kj}) = E(Z^4_{kji}|\theta_{kj}) \text{ and } \alpha_4(\mu_k) = E(\bar{z}^4_{kj}|\alpha_k).$$

We now state and prove the following Lemma.

Lemma 5.4.1 Assume conditions (i) and (ii) of Theorem 5.4.1. Assume in addition that

$$(a)\ E[\mu_4(\theta_{ij})] < \infty \text{ and } (b)\ E[\alpha_4(\mu_i)] < \infty.$$

Then, $V(MS_i) = O(K^{-1}) \qquad (i=1,2,3).$

Proof of the Lemma. Using the representation (5.3.11), and the fact that Z_{kji}'s are independent for different k, one gets by repeated application of c_R-inequality

$$V(SS_1) = E\sum_{k=1}^{K} V\{\sum_{j=1}^{n_k}\sum_{i=1}^{n_{kj}}(Z_{kji}-\bar{Z}_{kj})^2\}$$

$$\leq \sum_{k=1}^{K} E\{\sum_{j=1}^{m_k}\sum_{i=1}^{n_{kj}}(Z_{kji}-\bar{Z}_{kj})^2\}^2$$

$$= \sum_{k=1}^{K} n_k \sum_{j=1}^{m_k}\sum_{i=1}^{n_{kj}} E(Z_{kji}-\bar{Z}_{kj})^4$$

$$\leq 8\sum_{k=1}^{K} n_k \sum_{j=1}^{m_k}\sum_{i=1}^{n_{kj}} [E(Z_{kji}^4)+E(\bar{Z}_{kj}^4)]. \qquad (5.4.18)$$

Noting that

$$E(Z_{kji}^4) = EE[Z_{kji}^4|\theta_{kj}] = E[\mu_4(\theta_{11})]$$

$$E(\bar{Z}_{kj}^4) \leq n_{kj}^{-1}\sum_{i=1}^{n_{kj}} E(Z_{kji}^4) = E[\mu_4(\theta_{11})]$$

and using assumption (ii) of Theorem 5.4.1, one gets from (5.4.18)

$$V(SS_1) \leq [16\sum_{k=1}^{K} n_k^2]E[\mu_4(\theta_{11})] = 16C^2E[\mu_4(\theta_{11})] = O(K),$$

implying that $V(MS_1) = O(K^{-1})$.

Now, using the representation (5.3.12), and the fact that both $\bar{Z}_{kj}$'s and θ_{kj}'s are independent for diffent k, one gets after repeated application of c_β-inequality

$$V(SS_2) = \sum_{k=1}^{K} V[\sum_{j=1}^{m_k} n_{kj}[(\bar{Z}_{kj}-\bar{Z}_k) + (\theta_{kj}-\bar{\theta}_k)]^2\}$$

$$\leq \sum_{k=1}^{K} E\{\sum_{j=1}^{m_k} n_{kj}[(\bar{Z}_{kj}-\bar{Z}_k) + (\theta_{kj}-\bar{\theta}_k)]^2\}^2$$

$$\leq \sum_{k=1}^{K} m_k \sum_{j=1}^{m_k} n_{kj}^2 E[(\bar{Z}_{kj}-\bar{Z}_k) + (\theta_{kj}-\bar{\theta}_k)]^4$$

$$\leq 8\sum_{k=1}^{K} m_k \sum_{j=1}^{m_k} n_{kj}^2 [E(\bar{Z}_{kj}-\bar{Z}_k)^4 + (\theta_{kj}-\bar{\theta}_k)^4]. \qquad (5.4.19)$$

Applying c_β-inequality, one gets

$$E(\bar{Z}_{kj}-\bar{Z}_k)^4 \leq 8\{E(\bar{Z}_{kj}^4) + E(\bar{Z}_k^4)\}$$

$$E(\theta_{kj}-\bar{\theta}_k)^4 \leq 8\{E(\theta_{kj}^4) + E(\bar{\theta}_k^4)\}.$$

Now

$$E(\bar{Z}_k^4) \leq n_k^{-1}\Sigma_{j=1}^{n_k} n_{kj} E(\bar{Z}_{kj}) \leq E[\mu_4(\theta_{11})]$$

$$E(\bar{\theta}_k^4) \leq n_k^{-1}\Sigma_{j=1}^{n_k} n_{kj} E(\theta_{kj}^4) \leq E[m_4(\mu_1)]$$

so that

$$E(\bar{Z}_{kj}-\bar{Z}_k)^4 \leq 16E[\mu_4(\theta_{11})] \tag{5.4.20}$$

$$E(\theta_{kj}-\bar{\theta}_k)^4 \leq 16E[m_4(\mu_1)]. \tag{5.4.21}$$

Combining (5.4.19), (5.4.20), (5.4.21) and using assumption (11) of Theorem 5.4.2, one gets

$$V(SS_2) \leq C_0 \cdot K\cdot\{E[\mu_4(\theta_{11})]+E[m_4(\mu_1)]\},$$

where C_0 is a generic constant not depending on K. Thus, $V(SS_2) = O(K^{-1})$ as $K \to \infty$.

Using the representation (5.3.13), and the fact that $V[U_1+U_2+U_3] \leq 3[V(U_1)+V(U_2)+V(U_3)]$ for any random variables U_1, U_2, and U_3, one gets

$$V(SS_3) \leq 3[V\{\Sigma_{k=1}^{K} n_k(\bar{Z}_k-\bar{\bar{Z}})^2\} + V\{\Sigma_{k=1}^{K} n_k(\bar{\theta}_k-\bar{\bar{\theta}})^2\}$$

$$+V\{2\Sigma_{k=1}^{K} n_k(\bar{Z}_k-\bar{\bar{Z}})(\bar{\theta}_k-\bar{\bar{\theta}})\}]. \tag{5.4.22}$$

Since $\bar{Z}_k$'s are independent with $E(\bar{Z}_k) = 0$, one gets by an

application of (2.3.2) of Lemma 2.3.1,

$$V\Big[\Sigma_{k=1}^{K} n_k(\bar{z}_k - \bar{\bar{z}})^2\Big]$$
$$= \Sigma_{k=1}^{K}\Big(n_k - \frac{n_k^2}{n}\Big)^2\big[E(\bar{z}_k^4) - \{E(\bar{z}_k^2)\}^2\big]$$
$$+ 2 \sum_{k\neq k'} \frac{n_k^2 n_{k'}^2}{n^2} E(\bar{z}_k^2)E(\bar{z}_{k'}^2). \tag{5.4.23}$$

Now, using $E(\bar{z}_k^4) \leq E[\mu_4(\theta_{11})]$ and $E(\bar{z}_k^2) \leq \tau^2$ for all $k=1,\ldots,K$, one gets from (5.4.23)

$$V\Big\{\Sigma_{k=1}^{K} n_k(\bar{z}_k - \bar{\bar{z}})^2\Big\}$$
$$\leq \Sigma_{k=1}^{K}\Big(n_k - \frac{n_k^2}{n}\Big)^2 E[\mu_4(\theta_{11})]$$
$$+ 2 \sum_{k\neq k'} \frac{n_k^2 n_{k'}^2}{n^2} \tau^4. \tag{5.4.24}$$

Now, using assumption (ii) of Theorem 5.4.1, one gets from (5.4.23),

$$V\Big\{\Sigma_{k=1}^{K} n_k(\bar{z}_k - \bar{\bar{z}})^2\Big\} = O(K) \text{ as } K \to \infty. \tag{5.4.25}$$

By similar arguments, one has

$$V\Big\{\Sigma_{k=1}^{K} n_k(\bar{\theta}_k - \bar{\bar{\theta}})^2\Big\} = O(K) \text{ as } K \to \infty. \tag{5.4.26}$$

Using the formula $V(X) = EV(X|Y) + VE(X|Y)$, the c_r-inequality and writing $g(\theta_k) = \Sigma_{j=1}^{m_k} \mu_2(\theta_{kj})$, one gets

$$V\Big\{\Sigma_{k=1}^{K} n_k(\bar{z}_k - \bar{\bar{z}})(\bar{\theta}_k - \bar{\bar{\theta}})\Big\}$$

$$= V\{\Sigma_{k=1}^{K} n_k \bar{Z}_k (\bar{\theta}_k - \bar{\theta})\}$$

$$= E\{V[\Sigma_{k=1}^{K} n_k (\bar{\theta}_k - \bar{\theta}) \bar{Z}_k | \theta_{kj}\text{'s}]\}$$

$$= E\{\Sigma_{k=1}^{K} n_k^2 (\bar{\theta}_k - \bar{\theta})^2 n_k^{-2} \Sigma_{j=1}^{n_k} n_{kj} \mu_2(\theta_{kj})\}$$

$$= E\{\Sigma_{k=1}^{K} (\bar{\theta}_k - \bar{\theta})^2 \Sigma_{j=1}^{n_k} n_{kj} \mu_2(\theta_{kj})\}$$

$$\leq 2C\Sigma_{k=1}^{K} E\{(\bar{\theta}_k^2 - \bar{\theta})^2 g(\theta_k)\}$$

$$\leq 2C\Sigma_{k=1}^{K} E\{(\bar{\theta}_k^2 + \bar{\theta}^2) g(\theta_k)\}$$

$$\leq 2C\Sigma_{k=1}^{K} E^{1/2}(\bar{\theta}_k^2 + \bar{\theta}^2)^2 E^{1/2} g^2(\theta_k)$$

$$\leq 2C\Sigma_{k=1}^{K} E^{1/2}(2\bar{\theta}_k^4 + 2\bar{\theta}^4) E^{1/2}(m_k \Sigma_{j=1}^{m_k} \mu_2^2(\theta_{kj}))$$

$$\leq C_0 . K . \{E(\bar{\theta}_k^4)\}^{1/2} E^{1/2}(\mu_2^2(\theta_{11})) = O(K). \qquad (5.4.27)$$

Combining (5.4.25), (5.4.26) and (5.4.27) one gets $V(SS_3) = O(K)$ so that $V(MS_3) = O(K^{-1})$. The proof of the Lemma is now complete.

From Lemma 5.4.1, it follows that as $K \to \infty$, MS_i is a consistent estimator of $E(MS_i)$ $(i=1,2,3)$. This in turn implies that $\hat{R}_i \xrightarrow{p} R_i$ $(i=1,2)$ as $K \to \infty$ (Recall the definition of MS_i's in (5.3.28) and $\hat{R}_1$ and $\hat{R}_2$ in (5.3.32) and (5.3.33)).

Next we compute RSL($\xi, \underset{\sim}{e}_{EB}, \underset{\sim}{e}_0$). From (5.3.35), it follows that

$$E\|\underset{\sim}{e}_{EB}-\underset{\sim}{e}_B\|^2 = \Sigma_{k=1}^{K} E[\{1-\Sigma_{j=1}^{m_k} r_{kj} f_{kj}(1-\hat{B}_{kj})\}(\hat{\bar{u}}_k-u_k) + \Sigma_{j=1}^{m_k} r_{kj} f_{kj}(B_{kj}-\hat{B}_{kj})(\bar{X}_{kj}-u_k)]^2. \tag{5.4.28}$$

RSL ($\xi, \underset{\sim}{e}_{EB}, \underset{\sim}{e}_0$) can now be derived from (5.4.28) and (5.4.11).

The final theorem of this Section provides the asymptotic behavior of RSL ($\xi, \underset{\sim}{e}_{EB}, \underset{\sim}{e}_0$) as $K \to \infty$. More specifically, the following theorem is proved.

Theorem 5.4.2 Assume (i)-(iii) of Theorem 5.4.1 as well as (a)-(b) of Lemma 5.4.1. Then,

$$RSL(\xi; \underset{\sim}{e}_{EB}, \underset{\sim}{e}_0) \to 0 \text{ as } K \to \infty. \tag{5.4.29}$$

Proof. In view of (5.4.17), it suffice to show that

$$K^{-1} E\|\underset{\sim}{e}_{EB}-\underset{\sim}{e}_0\|^2 \to 0 \text{ as } K \to \infty. \tag{5.4.30}$$

Note that (5.4.30) also ensures the asymptotic optimality of $\underset{\sim}{e}_{EB}$.

To prove (5.4.28), first use the inequality

$$\begin{aligned} & E\|\underset{\sim}{e}_{EB}-\underset{\sim}{e}_B\|^2 \\ & \le 2\Sigma_{k=1}^{K} E[(\hat{\bar{u}}_k-u_k)^2 + \{\Sigma_{j=1}^{m_k} r_{kj} f_{kj}(B_{kj}-\hat{B}_{kj})(\bar{X}_{kj}-u_k)\}^2] \\ & \le 2\Sigma_{k=1}^{K} E[(\hat{\bar{u}}_k-u_k)^2 + m_k r_{j=1}^{m_k}(B_{kj}-\hat{B}_{kj})^2(\bar{X}_{kj}-u_k)^2] \end{aligned} \tag{5.4.31}$$

We shall now show that

$$K^{-1}\Sigma_{k=1}^{K}E(\hat{\bar{R}}_k-\mu_k)^2 \to 0 \text{ as } K \to \infty \tag{5.4.32}$$

$$K^{-1}\Sigma_{k=1}^{K}E[\Sigma_{j=1}^{m_k}(\hat{R}_{kj}-R_{kj})^2(\bar{X}_{kj}-\mu_k)^2] \to 0 \text{ as } K \to \infty. \tag{5.4.33}$$

Since, by assumption (ii) of Theorem 5.4.1, $\sup_{1\le k<\infty} m_k < C$, (5.4.30) will follow from (5.4.31)-(5.4.33).

To prove (5.4.33), first use the inequality

$$\max_{1\le j\le m_k}|\hat{R}_{kj}-R_{kj}| = \max_{1\le j\le m_k} n_{kj}|\hat{R}_1-R_1|(\hat{R}_1+n_{kj})^{-1}(R_1+n_{kj})^{-1} \le |\hat{R}_1-R_1|. \tag{5.4.34}$$

Since $\hat{R}_1 \xrightarrow{P} R_1$ as $K \to \infty$, from (5.4.34), one gets

$$\max_{1\le k\le K}\max_{1\le j\le m_k}|\hat{R}_{kj}-R_{kj}| \xrightarrow{P} 0 \text{ as } K \to \infty. \tag{5.4.35}$$

Moreover, from the calculations used to derive (5.4.10),

$$K^{-1}\Sigma_{k=1}^{K}E[\Sigma_{j=1}^{m_k}(\bar{X}_{kj}-\mu_k)^2]$$

$$\le K^{-1}\Sigma_{k=1}^{K}\Sigma_{j=1}^{m_k}\sigma^2(1-R_{kj})^{-1} \le C\sigma^2(R_1+1), \tag{5.4.36}$$

which implies that $K^{-1}\Sigma_{k=1}^{K}\Sigma_{j=1}^{m_k}(\bar{X}_{kj}-\mu_k)^2 = O_p(1)$.

Thus,

$$K^{-1}\Sigma_{k=1}^{K}\Sigma_{j=1}^{m_k}(\hat{R}_{kj}-R_{kj})^2(\bar{X}_{kj}-\mu_k)^2 \xrightarrow{P} 0 \text{ as } K \to \infty. \tag{5.4.37}$$

Moreover, the expression given in the lhs of (5.4.37) is bounded above by $K^{-1}\Sigma_{k=1}^{K}Y_k$, where

$$Y_k = \Sigma_{j=1}^{m_k}(\bar{X}_{kj}-u_k)^2 \leq 2[\Sigma_{j=1}^{m_k}(\bar{X}_{kj}-v)^2 + m_k(u_k-v)^2]. \tag{5.4.38}$$

Note that

$$E[\Sigma_{j=1}^{m_k}(\bar{X}_{kj}-v)^2]^2 \leq m_k\Sigma_{j=1}^{m_k}E(\bar{X}_{kj}-v)^4 \leq C_1 \tag{5.4.39}$$

for some constant C_1 which does not depend on k by using assumption (ii) of Theorem 5.4.1 and (a) and (b) of Lemma 5.4.1. Similarly, $E(u_k-v)^4 \leq C_2$ which does not depend on k under the same assumptions. Then, it follows from (5.4.38) that

$$\sup_{K\geq 1}\ \sup_{1\leq k\leq K} E(Y_k^2) \leq 4(C_1+C^2C_2). \tag{5.4.40}$$

Hence, from (5.4.40),

$$\sup_{K\geq 1} E[K^{-1}\Sigma_{k=1}^{K}Y_k]^2 \leq \sup_{K\geq 1} E[K^{-1}\Sigma_{k=1}^{K}Y_k^2] \leq 4(C_1+C^2C_2), \tag{5.4.41}$$

which proves the uniform integrability of $K^{-1}\Sigma_{k=1}^{K}Y_k$. This completes the proof of (5.4.33).

To prove (5.4.32), first noting that $\hat{R}_2\hat{d}_k\Sigma_{j=1}^{m_k}(1-\hat{R}_{kj}) = 1-\hat{d}_k$, and writing

$$\hat{q}_{kj} = (1-\hat{R}_{kj})/\Sigma_{j=1}^{m_k}(1-\hat{R}_{kj}) \text{ and } q_{kj} = (1-R_{kj})/\Sigma_{j=1}^{m_k}(1-R_{kj}),$$

one gets after some simplifications

$$\hat{\hat{u}}_k - u_k = \hat{d}_k(\hat{v}-v) + (1-\hat{d}_k)\Sigma_{j=1}^{m_k}(\hat{q}_{kj}-q_{kj})(\bar{X}_{kj}-v)$$
$$+ (d_k-\hat{d}_k)\Sigma_{j=1}^{m_k}q_{kj}(\bar{X}_{kj}-v). \tag{5.4.42}$$

Hence,

$$E(\hat{\theta}_k - \theta_k)^2$$

$$\leq 3E\Big[(\hat{\nu}-\nu)^2 + \Big[\Sigma_{j=1}^{m_k}(\hat{q}_{kj}-q_{kj})(\bar{X}_{kj}-\nu)\Big]^2$$

$$+ (\hat{d}_k-\hat{d}_k)^2\Big[\Sigma_{j=1}^{m_k}\hat{q}_{kj}(\bar{X}_{kj}-\nu)\Big]^2\Big].$$

Note that $\hat{d}_k\Sigma_{j=1}^{m_k}(1-\hat{B}_{kj}) = \hat{R}_2^{-1}(1-\hat{d}_k)$. Also,

$$\hat{\nu}-\nu = \{\Sigma_{k=1}^{K}(1-\hat{d}_k)\}^{-1}\Sigma_{k=1}^{K}(1-\hat{d}_k)\Sigma_{j=1}^{m_k}\hat{q}_{kj}(\bar{X}_{kj}-\nu). \tag{5.4.44}$$

Next, using $n_{kj} \geq 1$ and $\Sigma_{j=1}^{m_k} n_{kj} = n_k \leq C$, one gets

$$\Sigma_{j=1}^{m_k}(1-\hat{B}_{kj}) = \Sigma_{j=1}^{m_k} n_{kj}(\hat{R}_1+n_{kj})^{-1} \leq C(\hat{R}_1+1)^{-1}. \tag{5.4.45}$$

Moreover, we have already noted that $1-\hat{B}_{kj} \geq (\hat{R}_1+1)^{-1}$. Now,

$$1-\hat{d}_k = \hat{R}_2\Sigma_{j=1}^{m_k}(1-\hat{B}_{kj})/\{1+\hat{R}_2\Sigma_{j=1}^{m_k}(1-\hat{B}_{kj})\} \text{ and } 1\leq m_k\leq C.$$

Hence,

$$1-\hat{d}_k \geq \hat{R}_2 m_k(\hat{R}_1+1)^{-1}/\{1+\hat{R}_2 C(\hat{R}_1+1)^{-1}\} \geq \hat{R}_2/(\hat{R}_1+1+\hat{R}_2 C)\,; \tag{5.4.46}$$

$$1-\hat{d}_k \leq \hat{R}_2 C m_k(\hat{R}_1+1)^{-1}/\{1+\hat{R}_2 m_k(\hat{R}_1+1)^{-1}\}$$

$$= C\hat{R}_2 m_k/(\hat{R}_1+1+\hat{R}_2 m_k) \leq C(\hat{R}_2 C)/(\hat{R}_1+1+\hat{R}_2 C). \tag{5.4.47}$$

Combining (5.4.46) and (5.4.47), it follows from (5.4.44) that

$$|\hat{\nu}-\nu| \leq C^2 K^{-1}\Sigma_{k=1}^{K}\Big|\Sigma_{j=1}^{m_k}\hat{q}_{kj}(\bar{X}_{kj}-\nu)\Big|. \tag{5.4.48}$$

Write $Z_k = \left|\sum_{j=1}^{m_k} \hat{q}_{kj}(\bar{X}_{kj}-v)\right|^2 \le \sum_{j=1}^{m_k} \hat{q}_{kj}(\bar{X}_{kj}-v)^2$.

Hence, $Z_k^2 \le \sum_{j=1}^{m_k} \hat{q}_{kj}(\bar{X}_{kj}-v)^4 \le \sum_{j=1}^{m_k}(\bar{X}_{kj}-v)^4$. Under the given assumptions, it follows that $E(Z_k^2) \le C_0$, where C_0 does not depend on k. Now, $|\hat{v}-v|^2 \le c^4 K^{-1} \sum_{k=1}^{K} Z_k$ so that

$$\sup_{k\ge 1} E(\hat{v}-v)^4 \le c^4 \sup_{K\ge 1} \sup_{1\le k\le K} E(Z_k^2) \le c^4 C_0,$$

which implies the uniform integrability (in K) of $|\hat{v}-v|^2$. Besides, since

$$\max_{1\le k\le K} |\hat{d}_k - d_k| \le \left|\hat{B}_2 \sum_{j=1}^{m_k}(1-\hat{B}_{kj}) - B_2 \sum_{j=1}^{m_k}(1-B_{kj})\right|$$

$$\le |\hat{B}_2 - B_2| C + B_2 C \max_{1\le k\le K} \max_{1\le j\le k} |\hat{B}_{kj} - B_{kj}|, \tag{5.4.49}$$

using (5.4.35), one gets

$$\max_{1\le k\le K} |\hat{d}_k - d_k| \xrightarrow{P} 0 \text{ as } K \to \infty. \tag{5.4.50}$$

Using the above, and some more algebra, one gets $\hat{v} \xrightarrow{P} v$ as $K \to \infty$. Hence for proving (5.4.32), it remains only to show that

$$E[K^{-1} \sum_{k=1}^{K} \sum_{j=1}^{m_k} (\hat{q}_{kj} - q_{kj})(\bar{X}_{kj}-v)]^2 \to 0 \text{ as } K \to \infty, \tag{5.4.51}$$

and

$$E[K^{-1} \sum_{k=1}^{K} (d_k - \hat{d}_k)^2 [\sum_{j=1}^{m_k} q_{kj}(\bar{X}_{kj}-v)]^2 \to 0 \text{ as } K \to \infty. \tag{5.4.52}$$

This can be accomplished by showing that under our assumptions

(i) $$\max_{1\le k\le K}\max_{1\le j\le m_k}\left|\hat{\eta}_{kj}-\eta_{kj}\right| \xrightarrow{P} 0 \quad \text{as } K\to\infty, \tag{5.4.53}$$

(ii) $$\max_{1\le k\le K}\left|\hat{d}_k-d_k\right| \xrightarrow{P} 0 \text{ as } K\to\infty, \tag{5.4.54}$$

(iii) $$\{\Sigma_{j=1}^{m_k}(\hat{\eta}_{kj}-\eta_{kj})(\bar{X}_{kj}-\nu)\}^2$$

$$\le m_k\Sigma_{j=1}^{m_k}(\hat{\eta}_{kj}-\eta_{kj})^2(\bar{X}_{kj}-\nu)^2 \le CT_{j=1}^{m_k}(\bar{X}_{kj}-\nu)^2, \tag{5.4.55}$$

(iv) $$(d_k-\hat{d}_k)^2[\Sigma_{j=1}^{m_k}\eta_{kj}(\bar{X}_{kj}-\nu)\}^2 \le CT_{j=1}^{m_k}(\bar{X}_{kj}-\nu)^2, \tag{5.4.56}$$

and $\sup_{1\le k\le K} E[\Sigma_{j=1}^{m_k}(\bar{X}_{kj}-\nu)^2]^2 \le D$ which does not depend on K.

The details are omitted.

CHAPTER SIX
SUMMARY AND FUTURE RESEARCH

6.1 Summary

In summary, as noted in the previous sections, one can improve on Lee's estimators or predictors of several strata means or variances by incorporating information from similar neighboring sources. The proposed empirical Bayes estimators achieve this objective. The motivation of the different empirical Bayes estimators as well as the derivation of the different risk expressions does not require any specific distributional form of the prior other than the condition of posterior linearity. This indicates a clear robustness property of our estimation procedure. The Bayes risks of these empirical Bayes estimators are compared with the corresponding Bayes risks of the sample means or the sample variances indicating a clear edge for the former, especially for moderately large number of strata. On the basis of the numerical study undertaken in Chapter Two, one can say that, if there is indeed exchangeability, the empirical Bayes estimator outperforms the classical estimator, and does substantially better than the subjective Bayes estimator for misspecified priors. Thus, for exchangeable priors, the empirical Bayes estimator enjoys a great degree of robustness. The same phenomenon is expected to continue in other cases as well.

6.2 Future Research

In this dissertation, we have considered the problem of point estimation of different finite population parameters. We can search for a frequentist interpretation of our empirical Bayes estimators by suggesting some exact or approximate solution to the interval estimation problem. In this connection, the first task will be to find the mean squared errors of the proposed empirical Bayes estimators or some approximations to the same.

We can also extend our model by considering the multivariate case where we have several characteristics of interest. We can also explore the situation when we have more than one auxiliary characteristic available. The situation when we have response bias can also be explored.

BIBLIOGRAPHY

Basu, D. (1971). An essay on the logical foundations of survey sampling, part one. Foundations of Statistical Inference. Edited by V.P. Godambe and D.A. Sprott. Holt, Rinehart and Winston, Toronto, 203-242.

Bartoo, C.E. and Fuller, W.A. (1981). Prediction of county crop areas using survey and satellite Data. Proceedings of Section on Survey Research Methods. American Statistical Association Meeting, Detroit, MI, 500-505.

Cochran, W.G. (1939). The use of analysis of variance in enumeration by sampling. The Journal of the American Statistical Association, 34, 492-510.

Cochran, W.G. (1946). Relative accuracy of systematic and stratified random samples for a certain class of population. The Annals of Mathematical Statistics, 17, 164-177.

Cochran, W.G. (1977). Sampling Techniques. 3rd Edition. Wiley, New York.

Deely, J.J. and Lindley, D.V. (1981). Bayes empirical Bayes. The Journal of the American Statistical Association, 76, 833-841.

Diaconis, P. and Ylvisaker, D. (1979). Conjugate priors for exponential families. The Annals of Statistics, 7, 269-281.

Efron, B. and Morris, C. (1973). Stein's estimation rule and its competitors--an empirical Bayes approach. The Journal of the American Statistical Association, 68, 117-130.

Efron, B. and Morris, C. (1975). Data analysis using Stein's estimator and its generalizations. The Journal of the American Statistical Association, 70, 311-319.

Efron, B. and Morris, C. (1977). Stein's paradox in statistics. Scientific American, 236, 119-127.

Ericson, W. A. (1969a). Subjective Bayesian models in sampling finite populations (with discussions). The Journal of the Royal Statistical Society, B, 31, 195-233.

Ericson, W.A. (1969b). A note on the posterior mean. The Journal of the Royal Statistical Society, B, 31, 332-334.

Ericson, W.A. (1972). Subjective Bayesian models in sampling finite populations; stratification. Technical Report No. 15, Dept. of Statistics, University of Michigan, Ann Arbor, Michigan.

Ericson, W.A. (1976). A Bayesian approach to two-stage sampling. Technical Report No. AFFDL-TR-145, Dept. of Statistics, University of Michigan, Ann Arbor, Michigan.

Fay, R.E. and Herriot, R. (1979). Estimates of income for small places: an application of James-Stein procedures to census data. The Journal of the American Statistical Association, 74, 269-277.

Ferguson, T.S. (1973). A Bayesian analysis of some nonparametric problems. The Annals of Statistics, 1, 209-230.

Fuller, W.A. and Harter, R. (1985). The multivariate components of variance model for small area estimation. To appear in the Proceedings of the International Symposium on Small Area Estimation held in Carleton University, Ottawa.

Ghosh, M. and Meeden, G. (1986). Empirical Bayes estimation in finite population sampling. To appear in The Journal of the American Statistical Association, December issue.

Godambe, V.P. (1966). Bayes and empirical Bayes estimation in sampling finite populations. Technical Report No. 41, Dept. of Statistics, The John Hopkins University, Baltimore, Maryland.

Goldstein, M. (1975a). Approximate Bayes solution to some nonparametric problems. The Annals of Statistics, 3, 512-517.

Goldstein, M. (1975b). A note on some Bayesian nonparametric estimates. The Annals of Statistics, 3, 736-740.

Hartigan, J.A. (1969). Linear Bayes methods. The Journal of the Royal Statistical Society, B, 31, 446-454.

James, W. and Stein, C. (1961). Estimation with quadratic loss. Proceedings of the fourth Berkeley Symposium on Mathematical Statistics and Probability. University of California Press, Berkeley, 361-379.

Jewell, W. A. (1974a). Credible means are exact Bayesian for exponential families. Astin Bulletin, 8, 77-90.

Jewell, W.A. (1974b). Exact multidimensional credibility. Mitteilungen der Vereinigung Schweizerischer Versicherungsmathematiker, 74(2), 193-214.

Lindley, D.V. and Smith, A.F.M. (1972). Bayes estimates for linear model (with discussion). The Journal of the Royal Statistical Society, B, 34, 1-41.

Malec, D. and Sedransk, J. (1985). Bayesian inference for finite population parameters in multistage cluster sampling. The Journal of the American Statistical Association, 80, 897-902.

Morris, C. (1983a). Natural exponential families with quadratic variance functions: statistical theory. The Annals of Statistics, 11, 515-529.

Morris, C. (1983b). Parametric empirical Bayes inference: theory and applications. The Journal of the American Statistical Association, 78, 47-54.

Raj, D. (1968). Sampling Theory. McGraw-Hill, New York.

Rao, C.R. (1973). Linear Statistical Inference and Its Applications. 2nd Edition. Wiley, New York.

Rao, C.R. (1976). Characterization of prior distributions and solution to a compound decision problem. The Annals of Statistics, 4, 823-835.

Robbins, H. (1955). An empirical Bayes approach to statistics. Proceedings of the third Berkeley Symposium on Mathematical Statistics and Probability, Vol. Univetsity of California Press, Berkeley, 157-163.

Robbins, H. (1963). Some thoughts on empirical Bayes estimation. The Annals of Statistics, 11, 713-723.

Royall, R.M. (1976). The linear least squares prediction approach to two-stage sampling. The Journal of the American Statistical Association, 71, 657-664.

Royall, R.M. and Herson, J. (1973). Robust estimation in finite populations I. The Journal of the American Statistical Association, 68, 880-889.

Scott, A.J., Brewer, K.R.W. and Ho, E.W.H. (1978). Finite population sampling and robust estimation. The Journal of the American Statistical Association, 73, 359-361.

Scott, A. and Smith, T.M.F. (1969). Estimation in multistage surveys. The Journal of the American Statistical Association, 64, 830-840.

Sigman, R.S., Waruschak, G.A., Craig, M.E., Cook, P.W. and Cardenas, M. (1978). The use of regression estimation with LANDSAT and probability ground sample data. Survey Section Proceedings. American Statistical Association Annual Meeting, San Diego, California, 165-168.

Vardeman, S. and Meeden, G. (1983). Admissible estimators in finite population sampling employing various types of prior information. The Journal of Statistical Planning and Inference, 7, 329-341.

BIOGRAPHICAL SKETCH

Parthasarathi Lahiri was born on November 13, 1959, in Howrah, West Bengal, India. After graduating from Bansul C D P High School, West Bangal, India, in 1976, he enrolled at Presidency College in Calcutta. Upon receiving his Bachelor of Science degree with honors in statistics in 1979, he joined the Statistics Department of Calcutta University where he received a master's degree in statistics. After he received his degree in 1982, he joined R. K. Mission Residential College in India as a lecturer in statistics. He came to the United States in the Fall of 1983 when he joined the University of Florida at Gainesville. He expects to receive a degree of Doctor of Philosophy in statistics in August, 1986. During his time at the University of Florida, he was employed as a graduate consultant, performing consulting duties at the Institute of Food and Agricultural Sciences Statistics Unit. He was also employed as a teaching assistant in the Department of Statistics. Upon graduation, he will join the Department of Mathematics and Statistics, University of Nebraska at Lincoln.

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Malay Ghosh

Malay Ghosh, Chairman
Professor of Statistics

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Mark C. K. Yang
Professor of Statistics

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Ramon C. Littell
Professor of Statistics

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Kenneth M. Portier
Associate Professor of Statistics

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Michael A. DeLorenzo
Assistant Professor of Dairy Science

This dissertation was submitted to the Graduate Faculty of the Department of Statistics in the College of Liberal Arts and Sciences and to the Graduate School and was accepted for partial fulfillment of the requirements of the degree of Doctor of Philosophy.

August, 1986

Dean, Graduate School